AFOSR-TR-95

0568

Studies of the Effect of Image Degradation and Recombination

Final Report

March 1, 1992 -- September 1, 1995

AFOSR Grant # F49620-92-J-0176-P00001

PI -- William R. Uttal

Department of Industrial and Management Systems Engineering
Arizona State University
Tempe, AZ 85287-5906

Telephone and Fax -- (602) 965-8634
e-mail -- aowru@asu.edu

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION: NONE	1b. RESTRICTIVE MARKINGS: NONE
SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT: UNRESTRICTED
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S): PERLAB FINAL REPORT	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION: Dept. of IMSE - Arizona State Univ. — 6b. OFFICE SYMBOL (If applicable): 5906	7a. NAME OF MONITORING ORGANIZATION: AFOSR / NL
6c. ADDRESS (City, State and ZIP Code): Tempe, AZ 85287-5906	7b. ADDRESS (City, State and ZIP Code): 110 Duncan Ave, Suite B115 Bolling AFB DC 20332-8080
8a. NAME OF FUNDING/SPONSORING ORGANIZATION: AFOSR — 8b. OFFICE SYMBOL (If applicable): NL	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: F 49620-92-J-0176-P0003
8c. ADDRESS (City, State and ZIP Code): 110 Duncan Street, Suite B115 Bolling AFB, DC. 20332-0001	10. SOURCE OF FUNDING NOS. (see below)
11. TITLE (Include Security Classification): Studies of Effect of Image Degradation and Recombination	

10. SOURCE OF FUNDING NOS.

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT NO
61102F	2313	AS	
62205F	1123	00	

12. PERSONAL AUTHOR(S): Uttal, W.R.

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Yr., Mo., Day)	15. PAGE COUNT
Final	FROM 1Feb92 TO 1Sept95	1 September 1995	85

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB. GR.	

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

During the three year period of this grant, progress was made on two major series of experiments and some minor themes. In the first major one, we examined the effect of noise, brightness, contrast, and geometrical artifacts on a detection task simulating enhanced night vision devices. In the second, we explored the effects of noise, Fourier filtering, reduced acuity (by means of blocking) and combinations thereof on the discrimination and recognition of aircraft silhouettes and faces. The major empirical contribution of this work was the parametric exploration of a number of the key variables in visual perception. The major theoretical contribution was the proof that the Fourier components of an image were, at best, only a partial determinant of our perceptual response.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT: UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS ☐	21. ABSTRACT SECURITY CLASSIFICATION

22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE NUMBER (Include Area Code)	22c. OFFICE SYMBOL
John F. Tangney William R. Uttal	602-965-8634	NL

DD FORM 1473, 83 APR EDITION OF 1 JAN 73 IS OBSOLETE.

SECURITY CLASSIFICATION OF THIS PAGE

I Summary

During the three year period of this grant, progress was made on two major series of experiments and some minor themes. In the first major one, we examined the effect of noise, brightness, contrast, and geometrical artifacts on a detection task simulating enhanced night vision devices. In the second, we explored the effects of noise, Fourier filtering, reduced acuity (by means of blocking) and combinations thereof on the discrimination and recognition of aircraft silhouettes and faces. The major empirical contribution of this work was the parametric exploration of a number of the key variables in visual perception. The major theoretical contribution was the proof that the Fourier components of an image were, at best, only a partial determinant of our perceptual response.

II Research Objectives

1. This project carried out a program of research on the psychophysics of form. It specifically dealt with the perception of degraded and incomplete images. We were interested in the effect of image degradation on the ability of an observer to detect, discriminate, and recognize objects and scenes when the quality of the image has been reduced by systematic, quantified image transformations.

2. We were also interested in the complementary and closely related problem of sensory fusion -- how can multiple aspects or dimensions of the degraded and incomplete images be visually processed so that their subjective appearance is of a higher quality than it would otherwise be, and thus the observer's performance be enhanced from what it would otherwise be. The basic question in this complementary case deals with the ability of the visual system to integrate or combine low quality images of, for example, differing resolution to produce a high quality perception. It is a search for the rules of visual spatial combination and for the relative efficacy of what are distinguishable aspects or attributes of stimuli.

3. The two questions are, therefore, two aspects of the same problem. On the one hand, what effect does degradation have on the percept? and, on the other, how can the effects

of image degradation be overcome by utilizing the power of the visual system to integrate or combine degraded images?

4. We carried out a program of psychophysical studies that examined the effects of degradation and search on the nature of visual multidimensional combination. The long range goals of the psychophysical experiments were to provide information about the effects on performance of the human observer that go beyond rating or ranking of the subjective quality of an image.

5. We also carried out a program of computational modeling aimed at the simulation of human vision.

III Research Accomplishments

Two major series of experiments were carried out. The first dealt with a simulated analog of night vision devices. It used a detection task in which the images were degraded by the addition of random visual interference (visual noise), by varying the brightness, by varying the contrast, and by adding structured hexagonal artifacts. The work is completely described in detail in Uttal, Baruch, and Allen (1994). Our work provided the basic psychophysical foundations of vision using these devices. We were not able to uncover any previous work of this kind that dealt directly with night vision systems in the way we did.

In addition to the basic perceptual data, we discovered a curious learning effect. The size of the geometrical artifacts produced a curious pattern of responses in which the influence of intermediate size hexagons could be overcome by extensive experience but that of smaller hexagons could not. A third category, large hexagons, produced no effect from the outset.

The second major series of experiments dealt with degraded images in a more general way. Using both human faces and aircraft silhouettes, we studied the effect of combining well controlled image degradations. This first of several papers (Uttal, Baruch, and Allen, 1995a) employed a discrimination task in which small solid objects (aircraft silhouettes) were used as comparison targets. The subject's task was to determine if two sequentially presented objects were the same or different. In this series of experiments we used different combinations of three types of image degradations: acuity reducing averaging over variable sized blocks; Fourier low pass filtering with variable cutoff frequencies; and random visual interference. The ten experiments conducted indicated that the effect of Fourier filtering and blocking was generally small in all combinations and in all orders unless there was a substantial amount of visual interference present.

The next publication to report our work (Uttal, Baruch, and Allen, 1995b) dealt with recognition of these same aircraft silhouettes. In this case we used twelve different stimuli and required the subject to recognize rather than discriminate the stimuli. The most important theoretical development to emerge from this work has been the confirmation of the Harmon and Julesz phenomenon (i. e., low pass filtering of a blocked image enhances perception) for recognition but the rejection of it for discrimination. However, their theory cannot, in general, be correct since the phenomenon occurs for either order of degradation when the images are small.

Another secondary theme of our work dealt with the combination of degraded information. Experiments were carried out in which information from two different kinds of degraded (low-pass <u>filtered</u> and regionally averaged or <u>blocked</u>) visual stimuli were combined. In the first experiment, the degraded images were perceptually combined by being separately presented to each eye in a dichoptic viewing situation. Both stimuli were masked by identical random visual interference. When the two stimuli were dichoptically presented visually fused, performance in a discrimination task was enhanced over control situations in which only one of the two stimuli was presented. In the second experiment, the two degraded stimuli were physically superimposed prior to binocular presentation with a similar result. We concluded that a true advantageous information pooling was occurring when these two types of degraded stimuli were combined either physically or dichoptically. The implications of these findings for understanding the function of the visual system were discussed. This study (Uttal, Baruch, and Allen, 1995c) was also published.

Finally, we have also completed an extensive series of experiments studying recognition using faces as another model stimulus -- one that has an extensive previous history in studies like our own. The face is a stimulus that offers somewhat greater complexity and opportunities for examining a more realistically challenging domain. This project has resulted in a manuscript which has been submitted. (Uttal, Baruch, and Allen, Submitted.) It is included in this final report since it has not yet been published and it provides a good discussion of the general techniques and methods that we used in the Perception Laboratory.

In addition, we have also carried out a program of other research activities that have been closely related to the studies previously described. Many of these studies were in the field of computational modeling and dealt with the theory of combining image properties. This latter work, which was described in our original proposal, was also supported in part by the Office of Naval Research.

IV Publications

a. Published Articles:

1. Lovell, R., Uttal, W. R., Shepherd, T., & Dayanand, S. Texture segmentation of natural images using multiple weak operators and spatial averaging. Pattern Recognition, 1992, 25, 1157-1170.

2. Shepherd, T., Uttal, W. R., Dayanand, S., Lovell, R. A Method for shift, rotation, and scale invariant pattern recognition using the form and arrangement of pattern-specific features. Pattern Recognition, 1992, 25, 343-356.

3. Uttal, W. R. On models and mechanisms. Review of Newell, A. Unified Theories of Cognition. Behavioral and Brain Sciences. 1992, 15, 459-460.

4. Uttal, W. R. Toward a new behaviorism (1993). In Masin, S. (Ed.) Conceptual Foundations of Perceptual Theory. Amsterdam: North Holland.

5. Uttal, W. R. (1994). Pattern recognition. In Ramachandran, V. S. (Ed.) Encyclopedia of Human Behavior. Academic Press.

6. Uttal, W. R., Shepherd, T. Dayanand, S., Lovell, R. (1993). An integrated computational model of a perceptual-motor system. In Meyer, J., Roitblat, H. L., & Wilson, S. W., (Eds.) From Animals to Animats. Cambridge, MA: MIT Press.

7. Uttal, W. R. (1994). Introduction to Psychophysiology of Visual Masking (Bachmann, T. Author), Commack, New York: Nova Science Publishers.

8. Uttal, W. R., Davis, N. S., and Welke, C. (1994). Stereoscopic perception with brief exposures. Perception and Psychophysics, 56, 599-604.

9. Uttal, W. R., Baruch, T., & Allen, L. Psychophysical foundations of amplified night vision in target detection tasks. Human Factors, 1994, 36, 488-502.

10. Dayanand, S., Uttal, W. R., Shepherd, T., Lunskis, C. (1994). A particle system model for combining edge information from multiple segmentation modules. Computer Vision, Graphics, and Image Processing: Graphical Models and Image Processing. 56, 219-230.

11. Uttal, W. R. (1994) A new perspective on vision. (Review of Digital Images and Human Vision by Watson, A. B. (Ed.) Cambridge, MA: The MIT Press.) IEEE Multimedia, 1, 86-87.

12. Uttal, W. R. Baruch, T. & Allen, L. (1995a). The effect of combinations of image degradations in a discrimination task. Perception and Psychophysics, 57, 668-681.

13. Uttal, W. R. Baruch, T.,& Allen, L. (1995b). Combining image degradations in a recognition task. Perception and Psychophysics, 57. 683-691.

14. Uttal, W. R., Baruch, T. & Allen, L. (1995c) Dichoptic and physical information combination: A comparison. Perception, 24, 351-362.

15. Uttal, W. R. & Liu, N. An integrated vision system based on combining algorithms. Proceedings of the 18th International Conference on Computers and Industrial Engineering (ICC&IE'95). Shanghai, China. November 1995.

16. Uttal, W. R. Seeing by computing (1994). Arizona Infobahn. 1, 2-3.

17. Uttal, W. R., Liu, N.,& Kalki, J. An integrated computational model of three-dimensional vision. (Accepted for publication in Spatial Vision.)

18. Uttal, W. R., Baruch, T., & Allen, L. A parametric study of face recognition when image degradations are combined. (Submitted)

V Participating Staff

William R. Uttal, B. S, Ph.D.	PI
Todd Baruch, B. S.	Research Assistant
Linda Allen, B. S.	Graduate Student
Tom Shepherd, M. S.	Graduate Student
Jaggi Kalki, B. S.	Graduate Student
Robert Cole, Ph.D.	Visiting Scholar
Takeo Watanabe, Ph.D.	Assistant Professor

VI Coupling Activities

a. Presentations

Uttal,. W. R. 20 May 1992 -- Invited Talk -- Beijing Univ. PRC.

Uttal, W. R. 22 May 1992 -- Invited Talk -- Cheng Feng Science and Technology Group, Beijing, PRC.

Uttal, W. R. 24 July 1992 -- Paper presented at XXIst International Congress of Psychology, Brussels, Belgium.

Uttal, W. R. 6 December, 1992 -- Paper Presented at 2nd Annual Conference on the Simulation of Adaptive Systems, Honolulu, HI.

Uttal, W. R. 28 December 1992 -- Paper presented at William AFB.

Uttal, W. R. 17-22 January 1993. Interdisciplinary Conference, Jackson Hole, Wyoming.

Uttal, W. R. 14-15 May 1993. Society for Experimental Psychology, Seattle, Washington.

Uttal, W. R. 5-9th July 1993. Five invited lectures. University of Padua, Italy.

Uttal, W. R., Baruch, T., and Allen, L. 7 November 1993. The Psychonomics Society, Washington, D. C.

Uttal, W. R. et al. 9 November 1993. USN Symposium on Mine Countermeasures, Panama City Florida.

Uttal, W. R. July 1994. 3rd IEE International Workshop on Robot and Human Communication (RO-MAN'94). Nagoya, Japan.

Uttal, W. R. July, 1994. NTT Laboratories, Japan

Uttal, W. R. July, 1994. Tokyo University of Science

Uttal, W. R. July 1994. Riken Research Institute, Japan

Uttal, W. R. July 1994. Jikei University of Medicine

Uttal, W. R. 25 May 1995 ARVO, Ft. Lauderdale Fl

Uttal, W. R. 5 March 1995 Society of Experimental Psychologists

b. **Editorial**

Uttal was appointed consulting editor of the new APA journal Journal of Experimental Psychology: Applied in 1994.

VII New Developments

There were no patents submitted during the course of this study. Discoveries were of the kind described above.

VIII Possible Transitions

Our work is very relevant to the human factors involved in pilot performance. The face validity of our work on Night Vision Devices is obvious, but all of the work that we have done on degraded images is also important. In particular, we believe that our research on degraded image recognition would have provided some insight into the perceptual bases of the terrible tragedy in Iran in which "friendly fire" was responsible for the destruction of two US helicopters and their passengers and crews. Similarly, we believe that our work on modeling human visual perception of the kind we studied in this project will be very useful in the development of autonomous artificial vision systems.

IX. Appended Manuscript to Detail Methods and Theoretical Orientation

09/07/95 1:59 PM

A Parametric Study of Face Recognition when Image Degradations are Combined[1,2]

William R. Uttal, Todd Baruch, and Linda Allen
Perception Laboratory
Department of Industrial and Management Systems Engineering
Arizona State University
Tempe, AZ 85287-5906
Telephone and Fax -- (602) 965-8634
e-mail -- aowru@asuvm.inre.asu.edu

ABSTRACT

This article expands and quantifies one of the classic reports of modern visual perception research -- Harmon and Julesz' (1973) demonstration of an enhancement in recognition performance when area averaging (blocking) and spatial frequency filtering are sequentially applied. Our goals were twofold: First, to determine if the existence of the phenomena could be confirmed and replicated in a parametric study. Second, to determine if the new results supported the critical band masking theory originally

proposed by Harmon and Julesz. We confirmed the presence of the phenomenon for stimuli subtending approximately six deg of visual angle vertically, but observed a surprisingly different pattern of results for smaller stimuli subtending approximately one deg. These and other recent findings from other laboratories raise questions about their masking theory as a complete explanation of the phenomena.

INTRODUCTION

One of the most familiar and widely accepted phenomena of perceptual science is the demonstration by Harmon and Julesz (1973) that recognition can be improved by further degrading a local area averaged image by blurring it. The Harmon and Julesz demonstration was particularly interesting because the outcome was paradoxical -- the application of a second kind of degrading operator (spatial frequency filtering) seemed to counteract some of the decline in recognition performance produced by an earlier one (blocking). Despite the widespread popularity of this demonstration, to our knowledge the phenomenon has never been considered in a full blown parametric experiment. In this article, we expand and quantify their experimental design and determine under what conditions the phenomena may be considered to be generally valid.

This article builds upon results from our research program on the effects of sequentially combining image degradations on visual perception. In our earlier work (Uttal, Baruch, and Allen, 1995a; 1995b) we examined how cascading three different kinds of image degradations (spatial frequency "filtering", local area averaging -- "blocking," and visual interference) individually and collectively influenced discrimination and recognition, respectively, of stimulus forms that were initially solidly filled aircraft silhouettes. That is, prior to being degraded, the stimuli were uniformly white objects surrounded by uniformly black backgrounds. As the images were progressively processed by various combinations of degrading operators, internal regions took on a range of gray-scale values. Different degradations produced different effects. "Blocking" removed high frequency detail within the averaging blocks, but introduced spurious high frequency components at the edge of the blocks; low-pass spatial frequency filtering removed high frequency components, thus blurring the image; and visual interference produced a "fly-specked" appearance.

The results of these experiments raised questions about both the empirical and the theoretical foundations related to how we see degraded images, in general, and the Harmon and Julesz demonstration, in particular. Specifically, the findings of the first study (Uttal, Baruch, and Allen 1995a) were contrary to the expectations emerging from the pioneering study of cascaded degradations reported by them. Surprisingly, we showed that cascading image degradations in any order always produced a decline in <u>discrimination</u> performance. Discrimination to us was a same-different judgment made between two sequentially presented stimuli.

The second study (Uttal, Baruch, and Allen, 1995b) examined these same silhouette stimuli in a <u>recognition</u> task. Recognition to us was the assignment of a specific identifying number (1-12) to a single presented stimulus. The results were also unexpected. In this case, an enhancement of the participant's performance was observed comparable to that reported by Harmon and Julesz. However, this enhancement was not only obtained when the degradations were applied in the order they used (blocking followed by filtering), but also, to a somewhat lesser degree, when the two degradations were applied in the reverse order (filtering followed by blocking). According to their theoretical interpretation (high spatial frequency information masked the low frequency information necessary for recognition), this latter effect should not have occurred.

At the very least, therefore, we now know that there is a task dependence when processing sequentially degraded images -- the results for discrimination and recognition are qualitatively different. It is not certain, however, if this difference is due to different aspects of the stimulus used by the perceptual system in each task, whether it is the stimulus material itself, or if it is due to some other factor such as stimulus size.

Though we use faces as a stimulus in the present research, our main interest in them is as prototypical stimulus forms that vary in subtle, but psychophysically significant, ways to help us understand what happens when stimulus degradations are sequentially applied. For those interested in face recognition per se, we direct our readers to the important reviews of Ellis, Jeeves, Newcombe, and Young (1986), Bruce (1988), Young and Ellis (1989), and Bruce, Cowey, Ellis, and Perrett (1992), as well as the more recent experimental work of Bartlett and Searcy (1993).

Harmon and Julesz (1973) originally chose to frame their theory in terms of the spatial frequency components of the stimulus faces that they used in their demonstration. Others, (e.g., Tieger and Ganz, 1979; Riley and Costell, 1980; Zetsche and Caelli, 1989; Hayes, Morrone, and Burr, 1986, Costen, Parker, and Craw, 1994) have followed in this tradition by also emphasizing the spatial frequency spectral properties involved in the recognition of faces or other kinds of stimuli. An alternative approach is that the global perceptual organization or configuration aspects determine its recognizability.

The present report is essentially a replication, expansion, and quantification of the Harmon and Julesz (1973) demonstration. Our main purpose is to determine if their discovery of the paradoxical enhancement phenomenon occurs generally over a wider span of conditions than they used. We also wish to consider the generality and applicability of both their theory and a perceptual organizational approach in this broader context.

METHOD

Two series of experiments were conducted. The first series used relatively large (six deg) stimuli; the second used smaller (one deg) stimuli comparable in size to the silhouettes used in our earlier recognition studies.

EXPERIMENTAL SERIES 1 -- LARGE STIMULI

Participants

Well-trained, but naive about the purpose of the study, participants were used in each of the experiments reported in this article. Eight participants were used in Experiment 1, seven in Experiments 2 and 4, and six in Experiment 3. Each had normal or corrected vision and served in a series of five training sessions before participating in the experiment. The training sessions consisted of the same recognition protocol that would subsequently be used in the following experiments, but the stimuli were completely undegraded. All participants achieved recognition scores of 93% or greater for each of the faces presented in this manner. Participants were paid an hourly stipend and a bonus for completion of each experiment.

General Procedure

Four experiments were conducted in the first series using large stimuli. The first two were designed to provide base levels of recognition performance of the faces when only the spatial frequency filtering or the blocking was applied separately. Experiments 3 and 4 determined the effect of blocking and spatial frequency filtering when these degradations were combined in both orders; i. e., blocking followed by filtering, and filtering followed by blocking respectively.

The experimental procedure used in this study was fully automated. Participants signed into each session by typing their names on the computer keyboard. This initiated a sequence of actions in which the experiment assigned for that session was loaded and the computer configured to present the appropriate stimuli.

Participants were instructed to identify the single stimulus presented in each experimental trial by depressing the appropriate key on the top row of the computer keyboard. A master stimulus list consisting of photographs of the 12 faces used as stimuli was visible adjacent to the computer display throughout the experiment so that the participants could refer to it, if desired, prior to responding. A trial consisted of a sequence of visual displays on the CRT. The participant was first presented with a fixation stimulus consisting of four, dimly lit, outline corners of the viewing region. This was followed by a 500 msec blank period. One of the twelve faces was then presented for a nominal 100 msec. Following another 500 msec blank period, the four dim outline corners briefly appeared again on the display instructing the participant to respond. When the participant responded, the fixation corners for the next trial were displayed and the cycle repeated.

Within each experiment, all conditions were presented in a new random order each day to balance out any possible sequence effects. The stimulus conditions used for any trial was determined by random selection with replacement. We also included appropriate control conditions, as described later, for each experiment.

Stimuli

There are a number of differences between the faces used in this study and the single face of Lincoln that was used by Harmon and Julesz (1973). By using a collection of twelve very similar faces, rather than a single one, we were challenging the recognition performance of the participant much more severely. A single well recognized face, such as that of Abraham Lincoln, can be recognized by a highly distorted caricature emphasizing what often must be considered to be cognitively symbolic cues. It is possible that this caricature effect might have confounded the results of their demonstration. Another, and perhaps even more important difference, is that we have eliminated extraneous cues to recognition other than facial features per se. The stimulus faces were cropped by framing the face with a standard cutout template so that the details of hairstyle, hairline, and facial shape were removed. Only male faces of individuals with no beards or mustaches, unusual marks, or glasses were used. The set of twelve stimulus faces is shown in Figure 1. Each of the faces in this first series subtended a visual angle of approximately 3.5 deg (horizontal) x 6.05 deg (vertical).

INSERT FIGURE 1 HERE

Stimuli were processed by applying two forms of visual degradation (see our earlier works -- Uttal, Baruch, and Allen, 1995a and 1995b for a complete discussion of these degradations) -- (1) an "averaging over a region" algorithm (blocking) and/or (2) filtering the image to remove spatial frequencies higher than a prespecified nominal cut-off frequency from the spectrum of the Fourier transformed image. A main variable of the present study was the choice of applying a single one of these two degradations or the order in which both were applied.

Apparatus

The experiments were carried out on IBM PC compatible work stations with 486 Intel processors operating at 33 MHz. In this first series, participants were seated with their heads constrained by a chin rest so that their eyes were 84 cm from the face of the display. The entire experimental procedure was controlled by a computer program that randomly selected the stimuli, prepared the stimulus presentation sequence for each trial, collected the observer's responses, and then performed a preliminary analysis of the data obtained in each session. If a recognition error was made, auditory feedback of the correct answer was given by a computer speech generating system through earphones.

The CRT display itself (Tatung Model CM14SBS) was a raster scan, 34 cm (diagonal measurement) CRT with a full screen size of 1024 x 768 pixels. Its frame rate was 75 Hz and its horizontal refresh rate was 40 kHz. Since the active area (not the stimulus size) of the screen subtended 12.95 deg vertically and 16.17 deg horizontally, each pixel subtended 1.01 min vertically and .95 min horizontally. Given the uncertainty of the overlap of the adjacent point-spread-function from neighboring pixels on the phosphor, we consider the nominal width of a pixel to be 1 min of visual angle in the first series. It is, however, important to appreciate that the pixels are below resolution threshold in both series in this study.

The experimental room was indirectly lit by an incandescent bulb so that approximately one cd/m^2 fell on the screen as an ambient veiling light (Baker and Braddick, 1985; Farrell, Pavel, & Sperling, 1990; Groner, Groner, Muller, Bischof, & Di Lollo, 1993). The veiling light was measured by determining the amount of light reflected from a sheet of white paper at the surface of the display with a Tektronix J17 photometer equipped with a J1803 photometric sensing head.

The ambient veiling light also provided a constant lighting environment that stabilized the adaptation level, accommodation, and pupil size of the participants between trials. Since the exposure time of the stimulus was quite brief (100 msec), changes in all of these parameters were minimal or random from trial to trial. Since the stimuli were presented in random order in each experiment, intentional accommodative changes could not have been made. Therefore, no specific attempt was made to control pupil size or accommodation other than to maintain the visibility of the dimly lit display with the veiling light.

The contrast and intensity settings of the display were kept constant throughout both series in this study. When images were degraded, the image contrast changed in complex ways defined by the mathematics of the specific combination of degradation being applied. As we discussed in our earlier publications, image contrast cannot be designated by a single number such as RMS or peak-to-peak values. Any such single integrated number would ignore the pattern of bright and dark regions that defines each image. Rather, the specification of the degrading conditions, the initial stimulus, and the known properties of the display exactly define the intricate contrast conditions of a degraded stimulus.

As a general calibration procedure, the luminance of a test pattern consisting of a fully illuminated screen (i. e., all pixels set to white) was adjusted each day to 65 cd/m^2 with the veiling light present. Earlier work (Uttal, Baruch, and Allen, 1994) had shown that illumination levels and stimulus durations are not critical in this type of recognition experiment.

EXPERIMENTS AND RESULTS -- LARGE STIMULI

Experiment 1 -- The Solitary Effect of Spatial Frequency Filtering

Experiment 1 was designed to determine the solitary effect of spatial frequency filtering on the large face stimuli used in the first series of experiments. While baseline data of this kind had been obtained in our earlier recognition study (Uttal, Baruch, and Allen, 1995b) for stimuli of a different size and kind, those data obviously are not applicable for the larger face stimuli. The twelve large faces were, therefore, low-pass spatial frequency filtered by a non-ideal Butterworth filter as described in our earlier publications. Nominal cut-off frequencies of .43, .35, .26, and .17 cycles per degree of visual angle were selected as the values of this independent variable after pilot studies. The nominal cut-off of

.17 cycles/deg was the practical lower limit of our discrete spatial frequency filtering degradation. The next lowest step produced stimuli that were indistinguishable blurs. A control stimulus to which no filtering was applied was also inserted into the experimental design as a thirteenth randomly chosen alternative stimulus.

INSERT FIGURE 2 HERE

Figure 2 displays the results of Experiment 1. There is a progressive reduction in the recognition scores (measured as the percentage of the total number of presented stimuli that were correctly recognized) as the nominal cut-off limit of the filter is reduced. All data points in this figure and all subsequent ones have been plotted with standard error bars extending above and below the data symbol as measures of the variability of our results. Where the bars are not visible, they were smaller than the symbol for the plotted data value.

Since each daily session in this experiment consisted of approximately 325 trials and eight participants participated for five daily sessions, each point on Figure 2 represents a mean performance score based on 2600 trials.

Experiment 2 -- The Solitary Effect of Blocking

Experiment 2 was carried out to determine the solitary effect of blocking on the recognition of the large face stimuli. The sizes of the blocks used in this experiment (5, 10, 15, and 20 pixels on a side) were chosen in a pilot study. Block sizes larger than twenty pixels produced unrecognizable blurs.

INSERT FIGURE 3 HERE

The results of Experiment 2 are shown in Figure 3. There is a gradual reduction in recognition scores as the size of the block increases effectively reducing resolution. Since each daily session consisted of approximately 260 trials and seven participants participated for five daily sessions, each plotted point on Figure 3 represents a mean performance score based on 1820 trials.

Experiment 3 -- The Effect of Blocking Followed by Spatial Frequency Filtering

Next, in the first series, Experiment 3 reproduces the specific conditions of the original Harmon and Julesz (1973) demonstration -- blocking followed by spatial frequency filtering for a relatively large face. In this experiment, the large faces are first blocked at the three levels (10, 15, and 20 pixels on a side; values which are nominally equivalent to 10, 15, and 20 min of visual angle, respectively) used in the previous two experiments. Following the blocking, the stimulus faces are spatial frequency filtered at the three nominal cut-off frequencies (.35, .26, and .17 cycles/deg) used earlier.

INSERT FIGURE 4 HERE

In this and all of the remaining experiments, control conditions play a critical role. It is the set of controls against which the nine different experimental conditions must be compared <u>intra</u>experimentally to evaluate our results. In particular, in the remaining experiments, in which both kinds of degradation are sequentially applied to the stimulus, it is important to compare the results of the experimental conditions with results obtained when the stimuli were only degraded with the same values of blocking used in the experimental conditions -- but without either the preceding or following low-pass spatial frequency filtering degradation. In addition, three control stimuli in which the stimuli were only filtered were included in the design and are plotted with isolated points.

The results of Experiment 3 are shown in Figure 4. In this case, using large face stimuli, the enhanced recognition performance phenomenon occurs. That is, for the largest block size (except for the lowest cut-off filter) there is a small enhancement of the recognition scores for stimuli that have been blocked and then filtered, compared to the control condition of blocking alone. This improvement is in accordance with the classic paradoxical phenomenon originally described by Harmon and Julesz (1973). Though its magnitude is small when measured with a "percent correct" paradigm such as we have used here, it has a substantial perceptual effect.

Since each daily session consisted of approximately 325 trials and six participants participated for five daily sessions, each point in Figure 4 represents the mean performance score based on 650 trials.

The results of this experiment provide the first parametric measurement of the Harmon and Julesz (1973) demonstration. It is this, now validated and measured, yet seemingly inconsistent and paradoxical (in the sense that cascading degradations can, in some cases, improve performance) pattern of results that must be reconciled to understand exactly what is happening when degradations are combined. This reconciliation, however, requires that the generality of the phenomenon be examined. As we determined previously (Uttal, Baruch, and Allen, 1995b), a different pattern of results was obtained with small silhouettes -- both orders of the cascaded degradations produced the paradoxical enhancement in recognition performance. Because of this contradiction, it seemed prudent to determine if the difference was due to the stimulus size or the stimulus type. To accomplish this reconciliation, we must first carry out Experiment 4 to determine what happens when the order of the degradations is reversed for large faces.

Experiment 4 -- The Effect of Spatial Frequency Filtering Followed by Blocking.

Finally in the first series, Experiment 4 combined the two main forms of image degradation in the reverse order of Experiment 3 -- spatial frequency filtering followed by blocking. The motivation for this unusual experimental design came from our earlier results with small silhouettes (Uttal, Baruch, and Allen, 1995b). That study showed that an unexpected result -- the same paradoxical improvement in recognition -- occurred with this "reversed" order of degradations when using silhouette stimuli. As we shall see later in this article, when we examine the results of Experiment 8 which used small faces, the same surprising pattern results obtained in the earlier 1995b study also occurred there.

Because of the proliferation of conditions in our increasingly complex experimental designs we did not use all of the values of the block sizes and nominal cut-off filter frequencies that had been used in the earlier experiments. Only the three largest block sizes were used -- 10, 15, and 20 pixels on a side. Similarly only the three lowest nominal cut-off frequencies were used -- .17, .26, and .35 cycles/deg. The values were combined in all possible ways to produce nine different experimental conditions. Three other conditions were used in which only blocking was applied.

INSERT FIGURE 5 HERE

The results of Experiment 4 are shown in Figure 5. The major control condition, blocking alone, is shown with a solid line. Clearly the results of this experiment indicate that if a large stimulus face is low-pass filtered and then blocked (as indicated by the points connected by the broken lines) the effect is an overall reduction in performance compared to the stimuli that were only blocked. The performance levels for all of the experimental conditions in which both filtering and then blocking were applied (indicated by the three broken lines) are lower than the control conditions in which blocking alone was applied. This is in accord with the Harmon and Julesz (1973) explanation.

Since each daily session in this experiment had consisted of approximately 380 trials and seven participants participated for four daily sessions, each point on Figure 5 represents a mean performance score based on approximately 520 trials.

EXPERIMENTAL SERIES 2 -- SMALL STIMULI

The second series of experiments deals with small stimulus faces. It was, as noted, necessary to carry out these experiments because of the contradictions between the results of the first series in the present article, which used large faces, and the results of our earlier recognition study (Uttal, Baruch, and Allen, 1995b), which used small aircraft silhouettes. There are two possible explanations for these contradictory results. One is based on the stimulus type and the other is based on stimulus size. The following four experiments seek to unravel that contradiction by determining if it size or stimulus type that accounts for the difference in results for large and small stimuli.

General Method

In the second series of experiments in which small face stimuli were used, ten participants were used in Experiments 5 and 6 and nine participants were used in Experiments 7 and 8. All participants had normal or corrected vision and were extensively trained by participating in six preliminary training sessions. The first three training sessions used faces of varying sizes including approximately 5, 4, 3, 2, and 1 deg of visual angle measured vertically. These faces were undegraded. Performance was virtually perfect for all stimuli regardless of size.

The second three training sessions used only the small (one deg) face stimuli. These stimuli were also undegraded. All twelve faces were recognized nearly perfectly with little variation between them. As in the first series, participants were paid an hourly stipend and a bonus for completing the entire series.

Four experiments were conducted in the second series using the small (one deg) face stimuli. Except for the stimulus size and the dimensions of the blocks and the nominal low-pass cut-off frequencies of the filters, all other procedural elements of the second series were the same as the first series.

The stimuli used in the second series of experiments were considerably smaller than those used in the first series. The same twelve faces were reduced by computer manipulations and adjusting the viewing distance to subtend 0.75 deg (horizontal) and 1.0 (vertical). In this case, the eyes were 63.5 cm from the display. The viewing distance was also constrained by a chin rest. At this distance, a pixel nominally subtended 1.30 min vertically and 1.21 min horizontally. The nominal pixel in this case was, therefore, subtended about 1.25 min of visual angle, but again was unresolvable visually.

EXPERIMENTS AND RESULTS -- SMALL STIMULI

Experiment 5 -- The Solitary Effect of Spatial Frequency Filtering

Experiment 5, like Experiment 1, was designed to determine the solitary effect of spatial frequency filtering on the small face stimuli. The twelve faces were also low-pass filtered by a non-ideal Butterworth filter in the frequency domain, but in this case using nominal cut-off frequencies of 3.04, 2.6, 2.17, and 1.74 cycles/deg. These values were selected after pilot studies to bring our results into the middle range of approximately 30 to 80% recognition accuracy when the degradations were combined in Experiments 7 and 8. This experiment utilized only a single control condition -- the set of twelve unfiltered face stimuli.

INSERT FIGURE 6 HERE

The results of Experiment 5 are displayed in Figure 6. It can be seen that spatial filtering at these cut-off frequencies has a relatively modest effect on recognition of these small faces. This is similar to the results of Experiment 1. Since each daily session in Experiment 5 consisted of approximately 300 trials and ten participants participated for four daily sessions, each point on this curve represents a mean performance score based on 2400 trials. Standard error bars have been computed for all data points. If they are not visible, it indicates that the standard error is smaller than the symbol size.

Experiment 6 -- The Solitary Effect of Blocking

Experiment 6 provides the same calibration data for the small faces as did Experiment 2 for the large faces. That is, this experiment tracks the effect of block size on recognition performance. Because of the smaller overall size of the stimulus, smaller block sizes had to be used. On the basis of pilot studies, averaging block sizes of 2, 3, 4, 5, and 6 pixels were chosen. A control of the set of twelve undegraded faces was also used.

INSERT FIGURE 7 HERE

The results of Experiment 6 are shown in Figure 7. The block sizes used here produced a more profound effect on the small stimuli than did the larger blocks on the large stimuli. This is particularly notable since the blocks used on the small stimuli were relatively smaller than those used on the large stimuli and, therefore, should have, a priori, been expected to produce even a lesser effect than they did.

Since each daily session in this experiment consisted of approximately 260 trials and ten participants participated in 5 daily sessions, each point on this curve represents the average of about 2166 trials.

Experiment 7 -- The Effect of Blocking followed by Spatial Frequency Filtering.

Experiment 7 is the first in the second series to combine the two forms of degradation -- blocking and filtering -- and to determine their combined effect on the recognition of the small face stimuli. In this experiment, the conventional Harmon and Julesz order was followed -- blocking followed by spatial frequency filtering. The block sizes over which image intensities were averaged were 2, 3, 4, and 5 pixels respectively. The nominal cut-off frequencies of the Butterworth filter were 3.04, 2.60, 2.17, and 1.74 cycles/deg as in Experiment 5.

INSERT FIGURE 8 HERE

The results of Experiment 7 are plotted in Figure 8. Once again, the Harmon and Julesz phenomenon is replicated -- with the exception of a few of the smaller block sizes (four points at 2 pixels/block and one at 3 pixels/block) all of the stimulus conditions that were first blocked and then spatial frequency filtered are recognized better than when they were only blocked. Once again, cascaded degradations produce a paradoxical enhancement of the recognition scores.

Since each daily session consisted of approximately 450 trials and 9 participants participated for 6 daily sessions, each plotted point on Figure 8 represents the pooled results of approximately 1012 trials.

Experiment 8 -- The Effect of Spatial Frequency Filtering followed by Blocking.

Finally, Experiment 8 reverses the order of the degradations while maintaining all other conditions of Experiment 7. That is, the stimuli are first spatial frequency filtered at the four nominal cut-off limits used previously (3.04, 2.60, 2.17, and 1.74 cycles/deg) and then blocked with block sizes of 2, 3, 4, and 5 pixels.

INSERT FIGURE 9 HERE

The surprising effects of this "reversed" order of applying the degradations are shown in Figure 9. There is a paradoxical improvement in recognition for stimuli that are filtered and then blocked just as when the degradations were applied in the reverse order. This is the key finding in this study -- adding the high frequency information of the edges of the blocks to a previously blurred small face stimulus produces a more recognizable stimulus than one that had only been blocked. In other words, cascading these two types of image degradations, each of which reduces the information or image quality of that small face stimulus, in either order produces a smaller performance decrement than when only one degradation -- blocking -- is applied. This paradoxical effect found in this experiment cannot be explained by the removal of high frequency information by subsequent spatial frequency filtering as suggested by Harmon and Julesz (1973). High frequency spatial information is greater in the filtering-then-blocking conditions compared to when filtering only is applied. Furthermore, the effect is not just a quantitative change, but it is a complete qualitative reversal of the phenomenon when the results of small face stimuli and large face stimuli are compared.

This finding also resolves the type-versus-size uncertainty mentioned earlier. It is the size of the stimulus, not whether it was an aircraft silhouette or a face, that accounts for the qualitative difference between large and small stimuli in the comparable conditions when filtering is followed by blocking.

Since each daily session consisted of approximately 425 trials and nine participants each participated for six days, each plotted point in Figure 9 represents a mean performance score based on 956 trials.

Once again, we remind our readers that though the magnitude of the effects are relatively small in many of the experiments reported in this article, the standard errors are very small -- less than the 2% width of the symbols used in the figures. Therefore, the results are robust.

DISCUSSION

The results reported here and in our earlier articles emphasize the psychophysical complexity of the involved visual processes as conditions are added to the original demonstration described by Harmon and Julesz (1973). Our studies have only examined a few of the variables that affect how we recognize objects when degradations are combined. They do not speak to a host of other influences on recognition. Obviously, the shading and relative spatial position of facial features such as the eyes, nose, and mouth must be involved. Our interest has been directed at understanding the effects of the degrading operations with these other factors held constant or at least randomized. Obviously, the degrading operations will affect a host of different variables in different ways. The pattern of local and configurational changes are indisputably important, but we have not explicitly measured these independent variables. As we shall see, we believe that recognition is a redundant process with many cues sufficient to permit us to recognize a particular object.

Even more important at the outset of this discussion, it is important for our readers to understand that we make no pretense that we have a satisfactorily comprehensive theory to explain all of the findings from the many laboratories that have been or will be discussed. The exposition of a unified theory that combines all of the complex configurational, contrast, shading, and bandwidth properties of the stimulus is not possible at this time even though all of these factors are precisely defined by the image and the specific degradations that are applied. This article is patently aimed at disconfirming an older explanation that does not seem to have generality when the phenomenon under study is examined over a broader range of the involved parameters.

In order to emphasize how the variables that we have studied affect the outcome of this type of experiment, we now review the main findings from the three articles that make up this series so far.

1. In the first article (Uttal, Baruch, and Allen, 1995a), which used a discrimination paradigm and small silhouettes, no evidence of the paradoxical enhancement of performance reported by Harmon and Julesz (1973) was ever obtained. Cascading stimulus degradations always monotonically and progressively reduced discrimination performance. This non-paradoxical outcome was the qualitative, not just quantitative, opposite of the Harmon and Julesz (1973) demonstration for recognition. This is the key result of the 1995a study. Therefore, we concluded that there is a task dependency that produces qualitatively different results when blocking and low-pass filtering degradations are combined in discrimination and recognition, respectively. We have not yet determined whether this is due to the fact that the residual spatial frequency information following degradation of the stimulus does not determine the perceptual result or whether it is due to the fact that different attributes of a degraded stimulus are used in discrimination and recognition information processing, respectively.

2. In the second article (Uttal, Baruch and Allen, 1995b), we used a recognition task, but the same small (one deg of visual angle) silhouette stimuli used in

the discrimination study just described in 1. The paradoxical effect reported by Harmon and Julesz was obtained. However, contrary to their suggested explanation, the paradoxical increase in performance occurred when the blocking and filtering were sequentially applied in either order. This raised the question of whether it was the kind of stimulus material or the size of the stimuli that produced this difference.

3. The present study uses exactly the same recognition paradigm as in 2, but with face stimuli of two different sizes. In this case, the obtained pattern of results was as described by Harmon and Julesz for large stimuli (subtending approximately six deg). However, when small stimuli (subtending approximately one degree) were used, a surprising result was obtained: The paradoxical enhancement occurred with either order of combination of blocking and filtering just as it did with the one deg silhouettes.

There are other discrepancies that have to be explained before making a commitment to even a general theoretical approach to explaining these phenomena. Bachmann (1991) reports that there is an abrupt discontinuity at a particular block size (<18 pixels) that is not predicted by the "spatial frequency filtering model". (P.96). Parker and Costen (1993) in a brief abstract and then in a more comprehensive article (Costen, Parker and Craw, 1994) report much the same results but, to the contrary, interpret it as supportive of the spatial frequency approach that characterizes the Harmon and Julesz (1973) model.

It is important in establishing the context for the present discussions to appreciate that there is also still no agreement concerning which, if any, spatial frequency components of the stimulus are necessary for shape recognition. Ginsburg (1978), among others, suggested it is the low frequencies, while Fiorentini, Maffei, and Sandini (1983) argue for high frequencies. Hayes, Morrone, and Burr (1986), contrary to both of these reports, suggested that the spatial frequencies necessary for recognition reside in a band close to 20 cycles/face-width.

Costen, Parker, and Carew (1994) review the extensive literature on this problem and discuss the several different approaches that have been used to determine if there is an essential band of spatial frequencies necessary for face recognition. There is, as they note, considerable apparent disagreement among the many studies they cite, as well as those mentioned above, concerning an answer to the posed question. This is not surprising given the obviously redundant information carried by several different parts of the spatial frequency spectrum. Caricatures composed of a few lines and blurred renditions are all recognizable on the basis of different kinds of cues.

Following their review of the literature, Costen, Parker, and Craw (1994) come to the conclusion that:

> "Although these results show considerable variation and exhibit many ambiguities, the general conclusion could be drawn that there is a disproportionate decline in the accuracy of recognition of faces when the medium-low frequencies (approximately 8-16 cycles per face) are removed." (P. 130)

We also argue that, even if one is willing to accept the "ambiguities," such a conclusion would be justified only if the effects were constant over the size of the stimulus faces. The results of the present study indicate that there is, to the contrary, a qualitative difference between the results for large faces and small faces even though the cycles/face are approximately the same in Experiments 4 and 8. This suggests that the use of the normalized metric "cycles/face" may be inappropriate in this type of research. It also suggests that any conclusions concerning the masking effects of a "critical" band of spatial frequencies for face recognition may be specific to a certain size stimulus and may not be generalizable over the range of possible stimulus sizes.

It may be that this apparent disagreement is not real. Rather, as we suggested earlier, it may be that all of the findings are valid and complementary. Rather than being truly antagonistic, they may reflect the function of a number of alternative mechanisms for recognizing forms. It may be, therefore, that no single, universal theory of form recognition is possible.

We have demonstrated in this series of articles that there are task (discrimination results differ from recognition results), size (large images produce a qualitatively different pattern of results than do small ones), and visual interference effects (random dotted visual interference sometimes enhances and sometimes diminishes recognition performance) that must be taken into account. This brings us back to the original questions asked in our introduction. 1) Does the paradoxical result obtained by Harmon and Julesz (1973) generalize to other conditions? and 2) Does their model of high frequency masking of low frequencies satisfactorily explain these findings.

With regard to first question, the answer is two-fold. On the one hand, the paradoxical enhancement is found over a wide range of other parameter values. On the other hand, it is too general, it also appears in an unexpected place -- when small faces are first filtered and then blocked.

The answer to the first question provides the answer to the second question. Harmon and Julesz' explanation does not predict the results for the small face stimuli when blocking follows filtering. There appears to us to be no way that it can be modified to account for these new data because of the very strong premise it incorporates (high frequencies mask low frequencies).

What kind of theory can account for this pattern of results? While no specific answers are available for many of the questions generated by this line of research, there are two main alternative theoretical approaches to explaining our findings and those of other workers. The recognition of a face, or for that matter any object, could be based on the global or configurational aspects of the stimulus as opposed to the distribution and interaction of their spatial frequency components. However, efforts to apply organizational ideas are perpetually inhibited by the absence of satisfactory formal mathematical procedures to quantify exactly what it is that we mean by "configuration" or an "arrangement".

A review (Valentine, 1991) of the recognition literature and several recent reports (Bartlett and Searcy, 1993; Tanaka and Farah, 1993) emphasizing the configurational point of view reveals a generally non-quantitative approach to the problem. Configuration oriented research on face recognition has generally attempted to define the nature of the cues, components, and features that are salient to the recognition process. Efforts are sometimes made to link components in the spatial frequency domain to the configuration of components in the x, y domain. In general, however, these links are unconvincing, again reminding us that there is no satisfactory method yet available for quantifying configuration in the spatial domain. This methodological difficulty, however, does not mean that configuration, in some general sense, may not be the most salient cue of all. It does, however, leave researchers interested in the configurational approach without a tool comparable to the spatial frequency analysis method.

We now specifically reconsider the classic Harmon and Julesz "high spatial frequencies mask low spatial frequencies" premise. We have suggested that the qualitative differences between the results for small and large stimuli in the present experiment speak against their explanation. Before we can validate this conjecture, it is necessary to deal with an important possible confound. Could different sensitivities at different spatial frequencies on the contrast sensitivity function (CSF) account for the different results that were obtained with different stimulus sizes? We do not believe that an explanation based on varying contrast sensitivities can be complete. There are several reasons for this conclusion. First, both the higher component frequencies of the small stimuli and the lower frequencies of the large faces are lower than the peak of the CSF (approximately 5 cycle/deg) and on a relatively flat portion of the CSF. Second, contrast sensitivity is relatively high for both of these ranges of spatial frequencies and block sizes. Thus, while some quantitative differences between the two functions might not have been surprising, there is no discontinuity or inflection that could account for the total qualitative change occurring between the two experiments (4 and 8) in which spatial frequencies were followed by blocking.

Next, assume that the "standard response" is for the large faces and low spatial frequencies. Then consider that the small faces with their higher spatial frequencies fall into a region where the CSF is actually more sensitive than the region for the large faces. The implication is that the small face stimuli do not represent a less prototypical region of response (defined by CSF locus with a lower degree of sensitivity) than do the large faces. Rather, small faces exhibit a higher contrast sensitivity to the spatial frequencies of which they are composed. In this sense, it is the large faces that are less prototypical of the involved visual processes and mechanisms. Therefore, the results of Experiment 8 in the present study cannot be rejected on the basis of some low contrast sensitivity to the small stimuli. Rather, they should be considered to the prototype, not the large stimuli.

Now consider the research of Morrone, Burr, and Ross (1983) in which it was shown that adding high spatial frequency information (in the form of random "noise" filtered to approximate the band-width predicted by Harmon and Julesz to be the most inhibiting for recognition) actually increased the recognizability of the picture. Morrone and her colleagues also noted that the power of the high spatial frequency visual noise, as well as that of the edges produced by the blocking operation, was much less than the power of the low spatial frequency components. They assert that the "idea of the high-frequency components masking the low frequency components becomes less credible when their relative power is concerned." (P. 226).

Durgin and Proffitt (1993) provided another demonstration that challenges the Harmon and Julesz model. They first quantized a picture by blocking it to produce the usual degradation. Then, rather then filtering out the high frequency edge energy by low-pass filtering, they accentuated the high frequency power by laying a square grid of lines (with the same interline spacing as the width of the averaging block) over the blocked image. This essentially added power to the high frequency edges invoked by the Harmon and Julesz explanatory theory to mask the low frequency information necessary for recognition. However, the result was the opposite to that predicted by them -- the picture became more recognizable. It appeared, however, to be in back of a transparent screen. Durgin and Profitt suggest that this effect may be better described by the terminology of perceptual organization than as a result of any interaction between different spatial frequency components.

Subsequently, Morrone and Burr (1994) carried out a similar experiment in which they reversed the phase, without affecting the contrast, of the edges produced by the blocking operation. Blocked letter stimuli, for which recognition had been severely degraded by the blocking operation, also appeared to be seen behind a transparent screen as in the Durgin and Profitt demonstration. Equally surprising, their blocked stimuli were recognized as well as pristine, unblocked letters. Morrone and Burr (1994) attribute this finding to a process they describe as being dependent on the phase of the spatial frequency components (Morrone and

Burr (1988). They also suggest that it is difficult to account for their results in terms of the Harmon and Julesz masking model.

This brings us to the crux of the argument we wish to make. We suggest that the spatial frequency spectral properties of a stimulus object can be, at best, only a partial explanation of the complex of processes involved in face recognition, in particular, and form recognition, in general. We argue that the spatial arrangement of the parts of the face and how they are perceptually organized can be as influential, if not more influential, in recognition than the raw energetics of the spatial frequency components. This same general point is also made by Jenkins and Ross (1977), Meyer and Phillips (1980), and Watanabe (1995) for a different kind of visual phenomenon often attributed to early neural processing. All have shown that the perceptual organization of the scene can determine the nature of the McCullough (1965) effect. By doing so they also indicate that perceptual organization, presumably occurring at a relatively high level of the visual nervous system, can supplant or override the raw properties of the stimulus and, perhaps, even dominate visual mechanisms that are supposed to occur at a relatively low level in the nervous system.

Therefore, we suggest that frequency domain properties of an object are incomplete explanations of why we recognize objects. Turning an object upside-down, making a negative image of it, varying its familiarity, or jumbling the position of its components, on the other hand, can be totally disruptive to recognition even though the spatial frequency components may be roughly preserved. This suggests that the spatial domain properties of the stimulus may be more important than the frequency domain properties.

In conclusion, we asked two questions, one empirical and one theoretical. (1) Does the Harmon and Julesz (1973) phenomenon obtain over a wide range of stimulus parameters? and (2) Does the Harmon and Julesz theoretical explanation, based on the selective masking effect of a critical band of spatial frequencies, satisfactorily explain the effect?

The answer to the first question is affirmative. The paradoxical improvement of performance when a stimulus is first blocked and then spatial frequency filtered is present for at least the two quite different stimulus sizes used here. To our knowledge, the present article presents the first parametric and quantitative examination of this extremely popular demonstration. The magnitude of the phenomenon is relatively small, but it is highly robust given the small size of the standard errors. The major perplexity is that a paradoxical improvement in recognition performance can also occur in some cases if stimuli are first filtered and then blocked.

The second question cannot be completely answered at the present time. While it may be possible to modify the Harmon and Julesz explanation to account for all of the cited findings, several difficulties have arisen that suggest that it cannot currently be accepted as a completely satisfactory general model of degraded image recognition. Indeed, these new empirical findings raise the possibility that current modeling efforts may have been placed in entirely the wrong context. To summarize, these difficulties include:

1. For small faces, either order of degradation produces the phenomenon of enhanced recognition as shown in the present article.

2. The power of the high frequencies is insufficient to mask the low frequencies. (Morrone, Burr, and Ross, 1973)

3. Adding high frequency power in the form of a grid at the edges of the blocks does not degrade recognition performance. (Durgin and Profitt, 1993)

4. Reversing the phase of the edges of the blocks removes the inhibiting effect of the blocking operation (Morrone and Burr, 1994).

(Both groups in 3. and 4. report that the observer reorganizes the picture in a way that is equivalent to viewing the blocked object through a transparent grid or screen -- a kind of perceptual reorganization.)

5. Visual processing of the degraded images is task dependent. Therefore, it is not solely a function of the energy relations among the spatial frequency components of the stimulus. (Uttal, Baruch, and Allen, 1995a)

6. Random visual interference sometimes inhibits recognition (Uttal, Baruch, and Allen, 1995a and b) and sometimes enhances it (Morrone, Burr, and Ross, 1983).

7. There is strong evidence from several sources, including those mentioned earlier, that the perceptual organization of a scene can take priority over the raw

energetics and geometry of the stimulus in determining the perceptual response.

Obviously, a complete explanation of the paradoxical enhancement of recognition performance when a sequence of image degrading transforms is applied is not yet at hand. However, the difficulties mentioned above suggest that global, perceptual organizational concepts may also provide some insights into this complex human visual process that are not provided by theories based solely on the energetics of the Fourier components of an image.

In conclusion, what this complex of seemingly conflicting, redundant, and inconsistent results obtained from a variety of studies may be telling us is that there is no single answer to the question of how we recognize forms. It may be that none of the empirical observations may be "invalid" or "incorrect" -- they may each accurately assay a different mechanism. Each may be sufficient (but unnecessary) to account for the recognition process. From this point of view, there is no disagreement, only a varied set of different measures of a complex set of influential contributing perceptual processes. Clearly, the issue is more complicated than assumed previously and much additional research will have to be carried out to resolve the actual nature of the set of processes we collectively call "recognition."

REFERENCES

Bachmann, T. (1991). Identification of spatially quantized tachistoscopic images of faces: How many pixels does it take to carry identity? European Journal of Cognitive Psychology, 3, 87-103.

Baker, C. L. & Braddick, O. (1985). Temporal properties of the short-range process in apparent motion. Perception, 14, 181-192

Bartlett, J. C. & and Searcy, J. (1993). Inversion and configuration of faces. Cognitive Psychology, 25, 281-316.

Bruce, V. (1988). Recognizing Faces. Hillsdale, NJ: Erlbaum.

Bruce, V. Cowey, A. Ellis, A. W., & Perrett, D. I. (Eds.) (1992). Processing the Facial Image: Proceedings of a Discussion Meeting at the Royal Society of London. Oxford: Clarendon Press.

Costen, N. P., Parker, D. M., & Craw, I. (1994). Spatial content and spatial quantization effects in face recognition. Perception, 23, 129-146.

Durgin, F. H. & Profitt, D. R. (1993). Perceptual response to visual noise and display media. Final Report, NASA-NAG2-814. University of Virginia.

Ellis, H. D., Jeeves, M. A., Newcombe, F., & Young, A. (Eds.)(1986). Aspects of Face Processing. Dordrecht: Martinus Nijhoff.

Farrell, J. E., Pavel, M., & Sperling, G. (1990). The visible persistence of stimuli in stroboscopic motion. Vision Research, 30, 921-936.

Fiorentini, A., Maffei, L., & Sandini, G. (1983). The role of high spatial frequencies in face perception. Perception, 12, 195-201.

Ginsburg, A. P. (1978). Visual information processing based on spatial filters constrained by biological data. No. AMRL-TR-78-129. Aerospace Medical Research Laboratory.

Groner, R., Groner, M. T., Muller, P., Bischof, W. F., & Di Lollo, V.(1993). On the confounding effects of phosphor persistence in oscilloscopic displays. Vision Research. 33, 913-918.

Harmon, L. D. &. Julesz, B. (1973). Masking in visual recognition: Effects of two-dimensional filtered noise. Science, 180, 1194-1197.

Hayes, T., Morrone, M. C., & Burr, D. C. (1986). Recognition of positive and negative, bandpass-filtered images. Perception, 15, 595-602.

Jenkins, B. & Ross, J. (1977). McCollough effect depends upon perceived organization. Perception, 6, 106-154.

McCollough, C. (1965). Color adaptation of edge-detectors in the human visual system. Science, 149, 1115-1116.

Meyer, G. E. & Phillips, D. (1980). Faces, vases, subjective contours and the McCollough effect. Perception, 9, 603-606.

Morrone, M. C., Burr, D. C., & Ross, J. (1983). Added noise restores recognizability of coarse quantized images. Nature, 305, 226-228.

Morrone, M. C. & Burr, D. C. (1988). Feature detection in human vision: A phase dependent energy model. Proceedings of the Royal Society of London. Series B, 235, 221-245.

Morrone, M. C.& Burr, D. C. (1994) Visual capture and transparency in blocked images. Perception, 23, 20.

Parker, D. M. & Costen, N. P. (1993). Spatial frequency and internal masking in face images. Perception, 22, 48.

Riley, D. & Costell, A. (1980). Comment on "Recognition of faces in the presence of two-dimensional sinusoidal masks " by Tieger and Ganz. Perception and Psychophysics, 27, 373-374.

Tanaka, J. W. & Farah, M. J. (1993).Parts and wholes in face recognition. Quarterly Journal of Experimental Psychology, 46A, 225-245..

Tieger, T. & Ganz, L. (1979). Recognition of faces in the presence of two-dimensional sinusoidal masks. Perception and Psychophysics, 26, 163-167.

Uttal, W. R., Baruch, T., & Allen, L. (1994). Psychophysical foundations of a model of amplified night vision in target detection tasks. Human Factors, 36, 488-502.

Uttal, W. R., Baruch, T., & Allen, L. (1995a). The effect of combinations of image degradations in a discrimination task. Perception and Psychophysics, 57, 668-681.

Uttal, W. R., Baruch, T., & Allen, L. (1995b). Sequential image degradations in a recognition task. Perception and Psychophysics, 57, 682-691.

Valentine, T. (1991). Representation and Process in Face Recognition. In Watt, R. J. (Ed.) Pattern Recognition by Man and Machine. Boca Raton: CRC Press, Inc.

Watanabe, T. (1995). Orientation and color processing for partially occluded objects. Vision Research, 35, 647-655.

Young, A. W. & Ellis, H. D.(Eds.) (1989). A Handbook of Research on Face Processing. Amsterdam: Elsevier Science.

Zetsche, C. & Caelli, T. (1989). Invariant pattern recognition using multiple filter image representations. Computer Vision, Graphics, and Image Processing, 45, 251-262.

FIGURE CAPTIONS

1. The twelve faces used in this study. The faces were cropped as indicated to avoid the secondary cues of hair outline that might confound the results for the recognition task that was the target of this study. None of the faces had any distinctive feature such as a mustache, beard, or glasses. The differences in the faces are solely due to the nature and arrangements of the facial features.

2. The results of Experiment 1 for large faces in which base level recognition scores were measured as a function of the nominal low-pass cutoff frequency (measured in cycles/deg of visual angle -- c/d). The lower the nominal spatial frequency cut-off frequency, the more degraded (blurred) were the stimulus faces. "CNT" indicates a control stimulus in which no degradation was applied. The vertical axis in this figure and all subsequent ones indicates the percentage of presented stimulus faces that were correctly recognized. In this figure and all subsequent ones, the standard error is plotted as vertical whiskers on individual symbols. In most cases the standard error is smaller than the symbol and cannot be seen. This is the

result of the relatively large number of trials pooled in each experiment.

3. The results of Experiment 2 in which base level recognition scores for large faces were measured as a function of the block size. As the block size (measured in terms of the number of pixels along a side of the square averaging region) increases, recognition performance decreases. In this case "CNT" refers to the control score for all faces when no blocking was applied.

4. The results of Experiment 3 for the large face stimuli in which two forms of degradation were sequentially applied to large faces -- blocking was followed by filtering. The control stimuli in this figure are of two kinds. First, the three isolated symbols indicate the scores for the large face stimuli that were only low-pass filtered. The second control is plotted along the solid line and represents stimuli that had only been blocked. The points on the three broken lines are for the experimental stimuli that had been both blocked and then filtered. In this case, at the largest (most degraded) block sizes, spatial frequency filtering enhances recognition scores; but only for the .35 and .26 cycles/deg nominal cut-off

frequencies at the largest block size. This is the classic Harmon and Julesz paradoxical phenomenon.

5. The results of Experiment 4 in which two forms of degradation were sequentially applied to large faces. This is the same experimental design as Experiment 3 with the exception that the order of the filtering and blocking degradations are reversed -- the filtering was applied first and the blocking second. No evidence of any enhancement in recognition performance is indicated in these results.

6. The results of Experiment 5 for small faces in which the solitary effects of the nominal cut-off frequencies are measured. There is a gradual reduction in recognition performance as the nominal low-pass cut-off limit is lowered. Because the faces are smaller, the appropriate spatial frequencies are higher than in Experiment 1. It should be noted again that the standard error bars are usually not visible because they are smaller than the symbols used in plotting the graph.

7. The results of Experiment 6 for small faces in which the solitary effects of block size are measured. There is a gradual reduction in recognition as the size of the averaging block increases. Because of the small

stimulus size, the block sizes (measured as the size of the side of the square region over which the averaging occurs) must be smaller than in Experiment 2.

8. The results of Experiment 7 for small faces in which the combined effects of sequentially degrading stimuli by blocking and then low-pass spatial frequency filtering are measured. The paradoxical Harmon and Julesz phenomenon is shown to be present in this condition since the dotted lines (representing recognition performance for blocking followed by filtering) are in many cases higher than the solid line representing the recognition of stimuli that have only been blocked.

9. The results of Experiment 8 for small faces in which the combined effects of sequentially degrading stimuli by low pass filtering and then blocking are measured. Surprisingly, a paradoxical enhancement of recognition performance occurs in this case as well as under the conditions of Experiment 7. This outcome is contrary to the predictions of the Harmon and Julesz theory.

10. The Gamma function of the display used in this experiment. The binary values transmitted from the computer are plotted on the horizontal axis. The

luminance of the resulting signal is plotted on the vertical axis.

APPENDIX A -- Technical Comment: The Gamma Function

Another important consideration in the use of a Cathode Ray Tube(CRT) display in a study of this kind is the CRT's Gamma function -- the relation between the binary coded input and the resulting luminosity of the CRT display. We measured the Gamma function for our display adjusted to the conditions described in the experimental procedure section. This function is shown in Figure 10.

INSERT FIGURE 10 HERE

This curve shows that, at the display settings used in this study, the lowest binary values did not produce measurable light. Furthermore, the screen appeared to be totally dark at the lowest settings. This would cause the dimmest pixels in the stimulus images to all appear equally dark. However, it is important to note that this same Gamma function was used for both the large stimuli of Series 1 and the small stimuli of Series 2. Whatever effect this non-linearity at the low light levels had on the display is common to both and, therefore, could not have accounted for the qualitative differences in obtained results for the two series. Furthermore, while the luminosity of the display might have been distorted in this manner, the mathematics of the image degradations was perfectly linear, being carried out on the internal representations of the images in the computer memory, not on the images displayed on the CRT. This information, along with the original stimuli and the specification of the applied degradations given in the text, precisely defines each of the hundreds of stimulus images used in this study.

AUTHOR NOTES

1. This study was supported by Grant #F49620-92-J-0176 P0001 from the United States Air Force Office of Scientific Research.

2. We acknowledge with appreciation the comments made by our colleagues Takeo Watanabe and MaryLou Cheal on an early version of this manuscript.

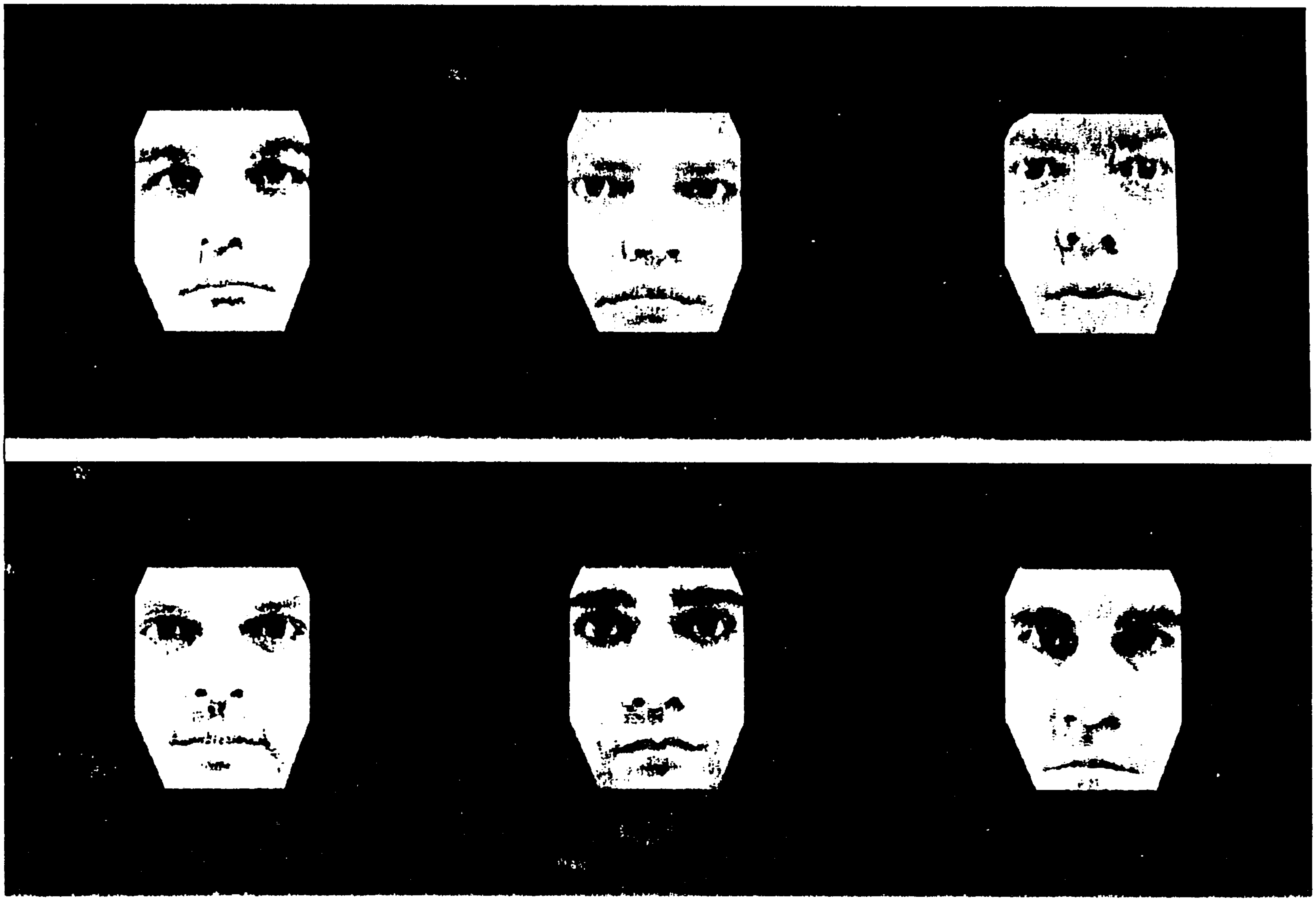


FIGURE 1 (First Part)

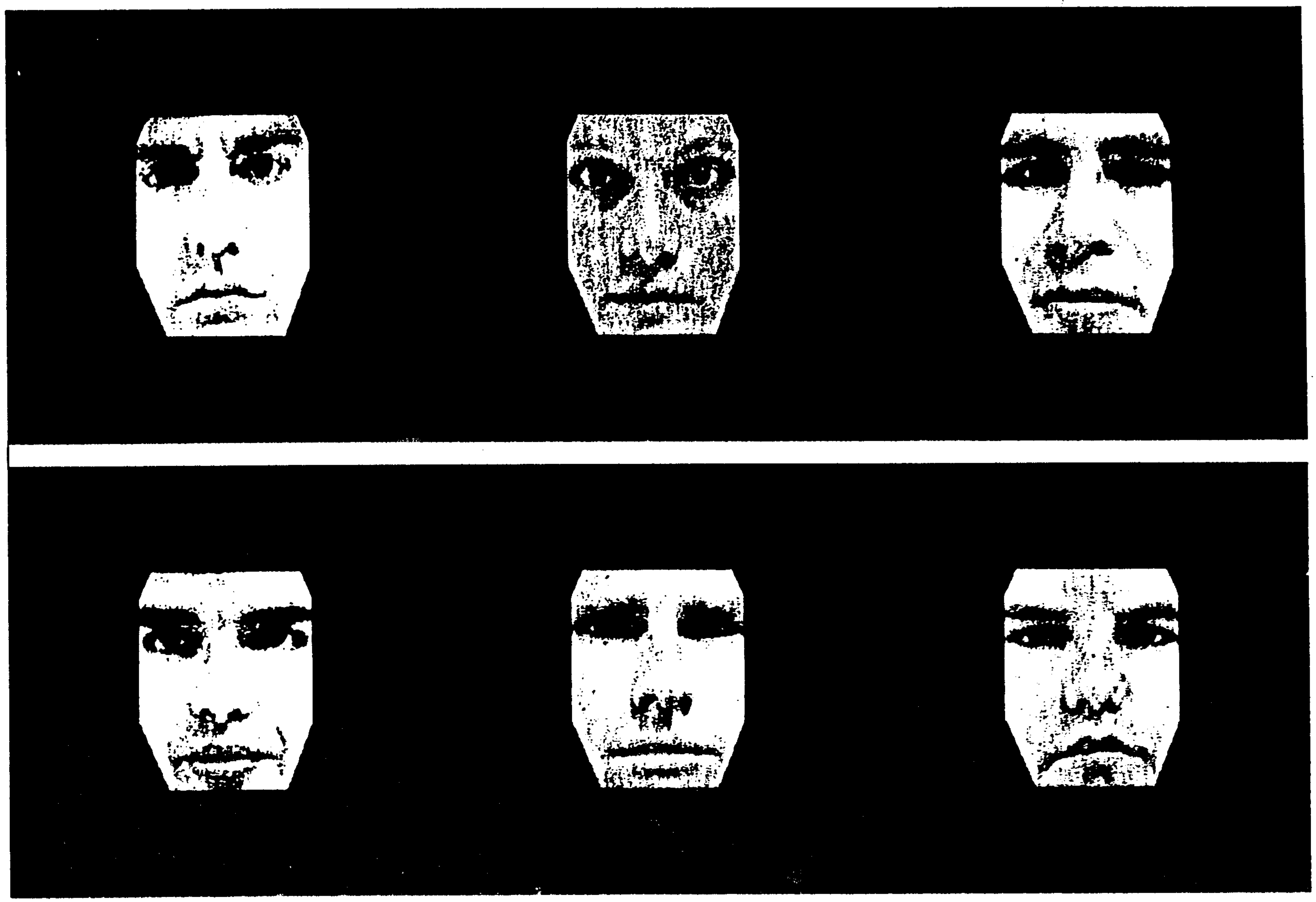


FIGURE 1 (Second Part)

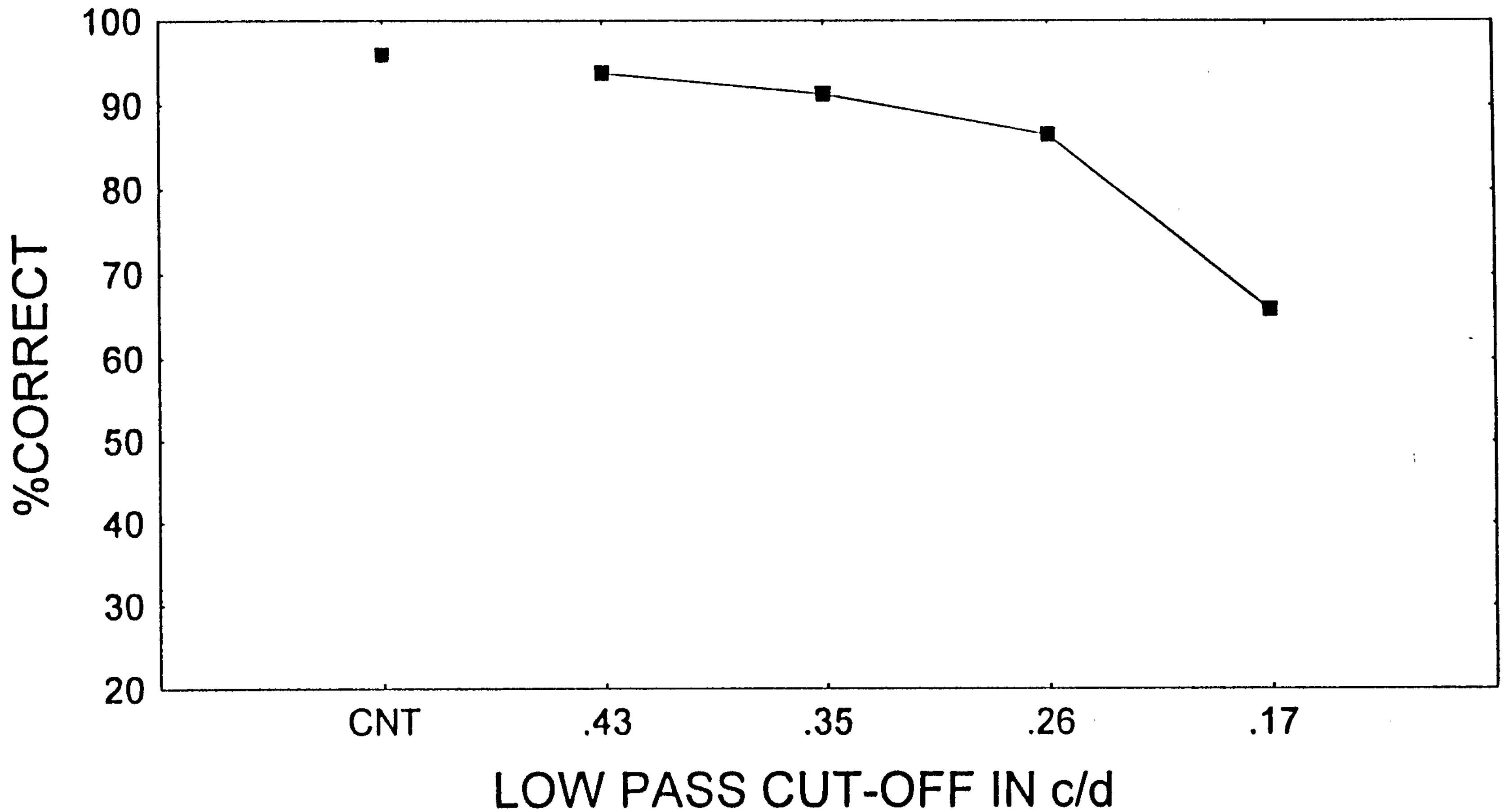

100
90
80
70
60
50
40
30
20
%CORRECT
CNT
.43
.35
.26
.17
LOW PASS CUT-OFF IN c/d

FIGURE 2

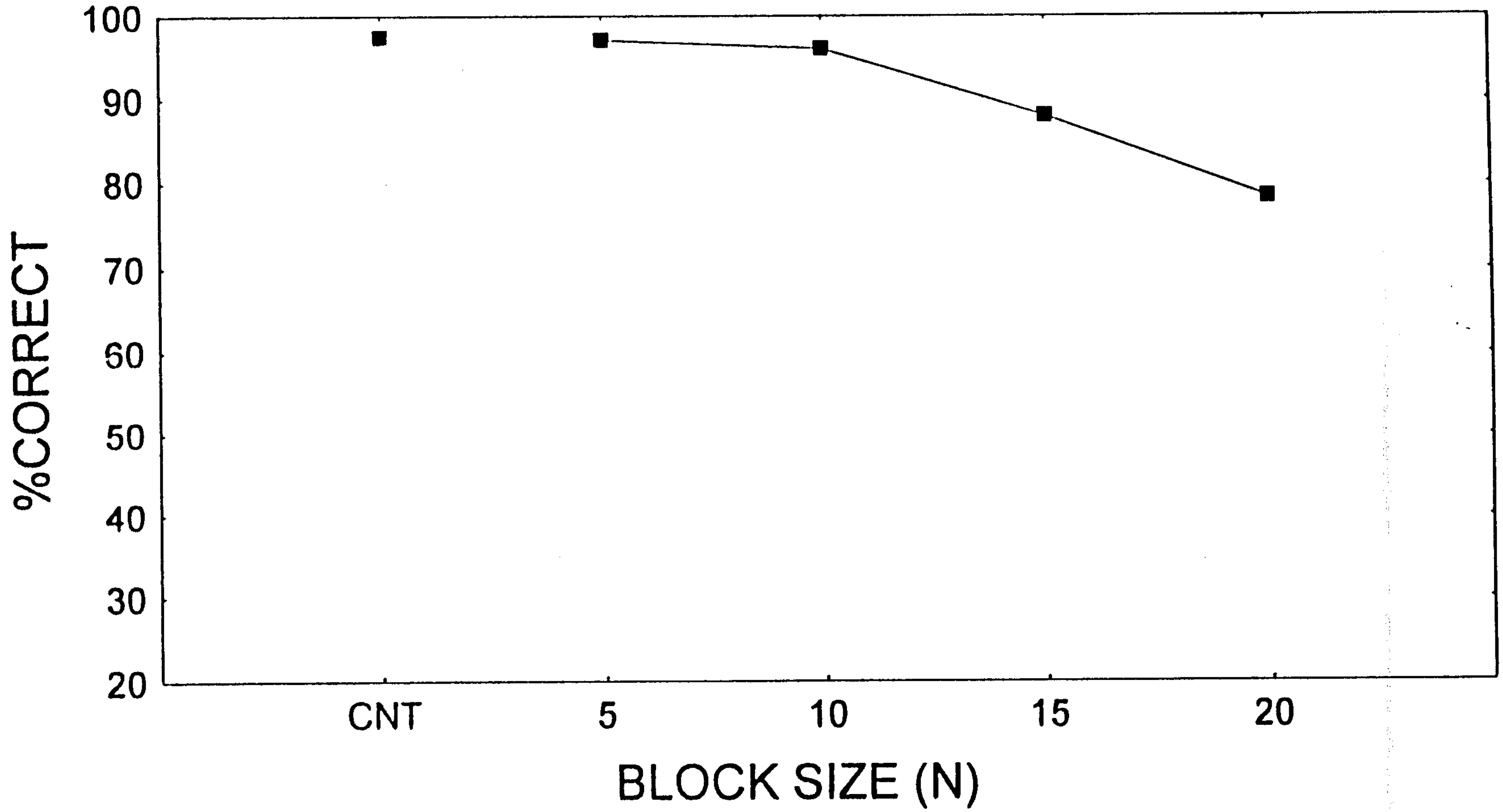


FIGURE 3

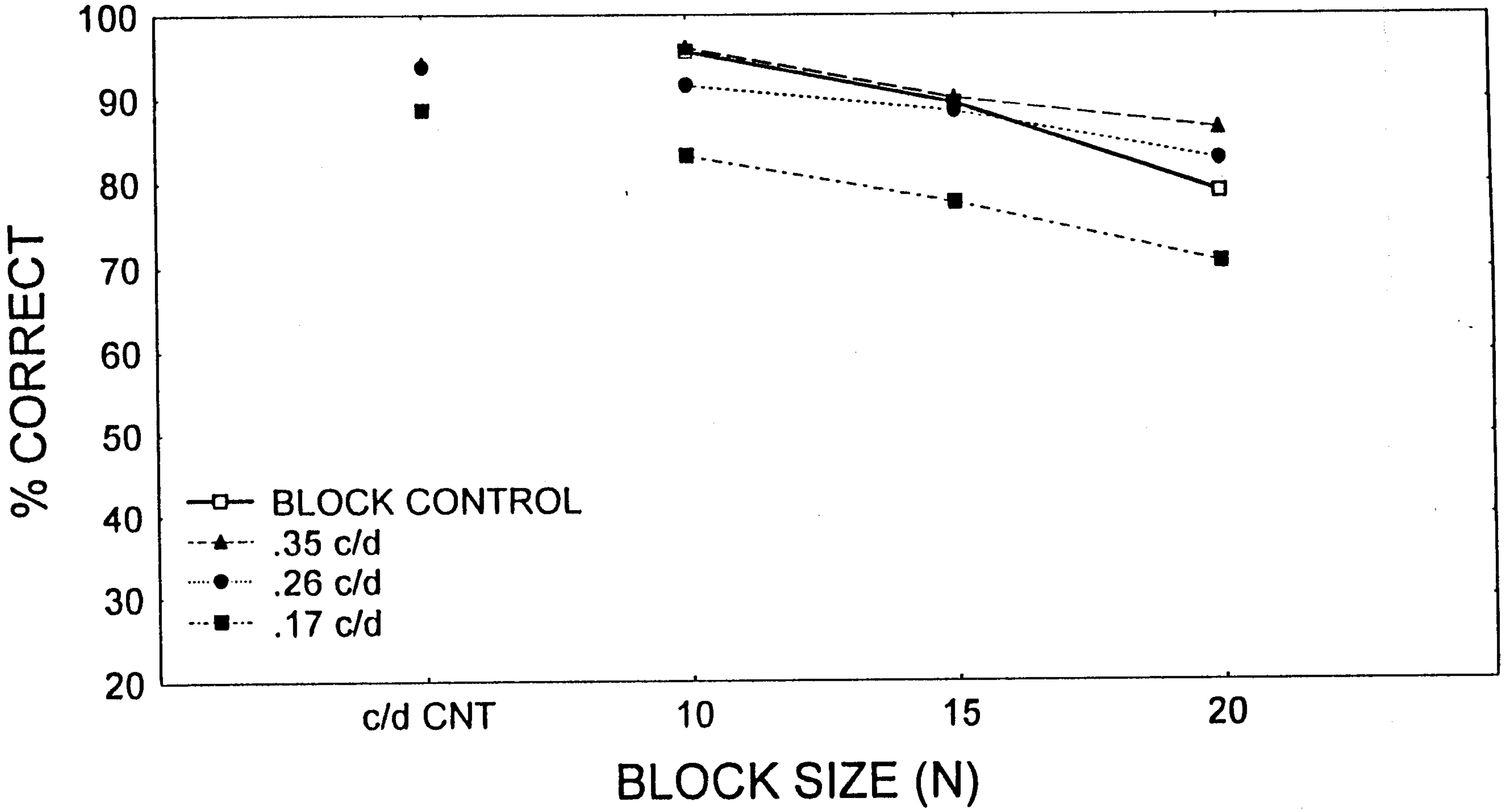
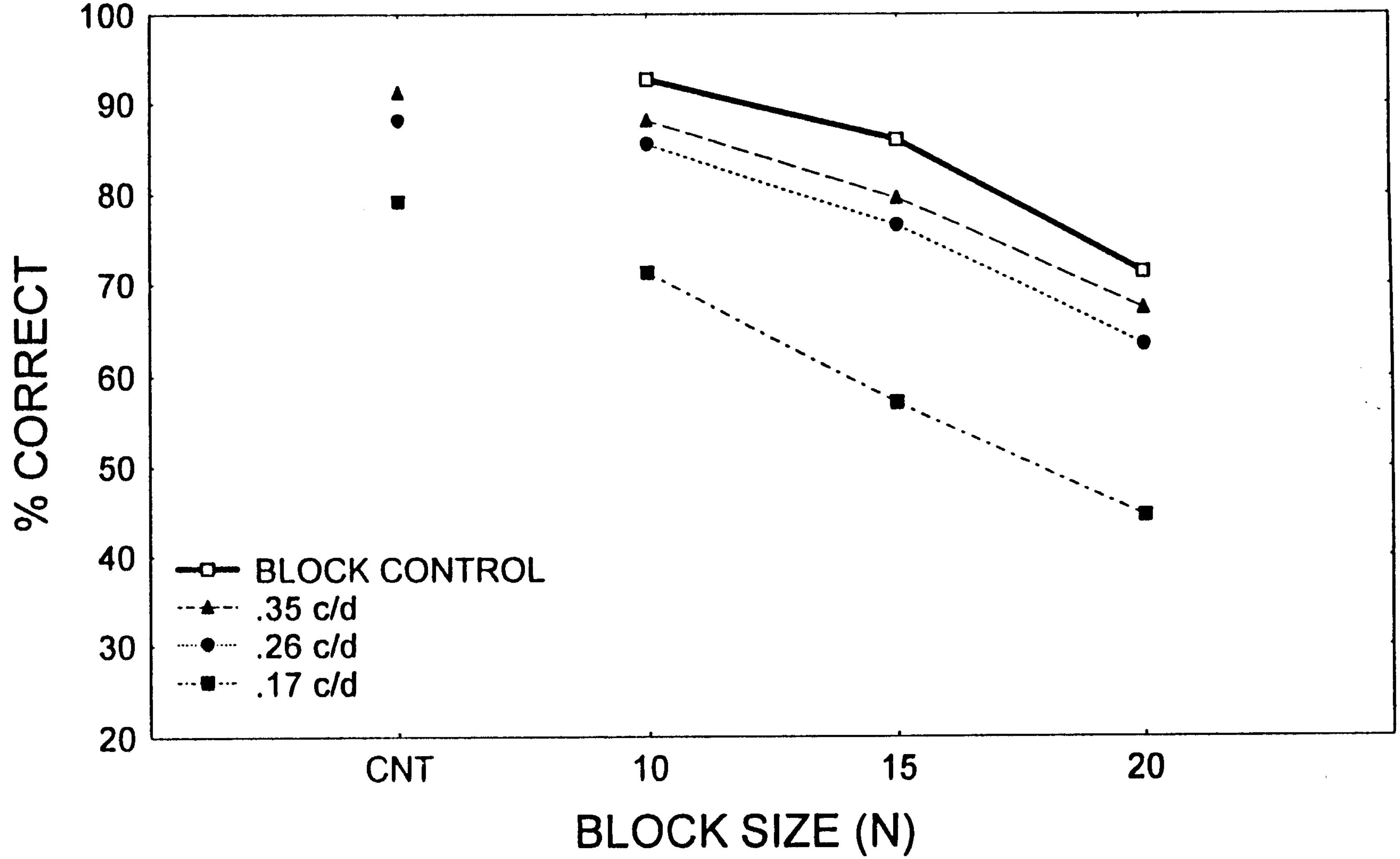

100
90
80
70
60
50
40
30
20
% CORRECT
BLOCK CONTROL
.35 c/d
.26 c/d
.17 c/d
c/d CNT
10
15
20
BLOCK SIZE (N)

FIGURE 4



100
90
80
70
60
50
40
30
20
% CORRECT
BLOCK CONTROL
.35 c/d
.26 c/d
.17 c/d
CNT
10
15
20
BLOCK SIZE (N)

FIGURE 5

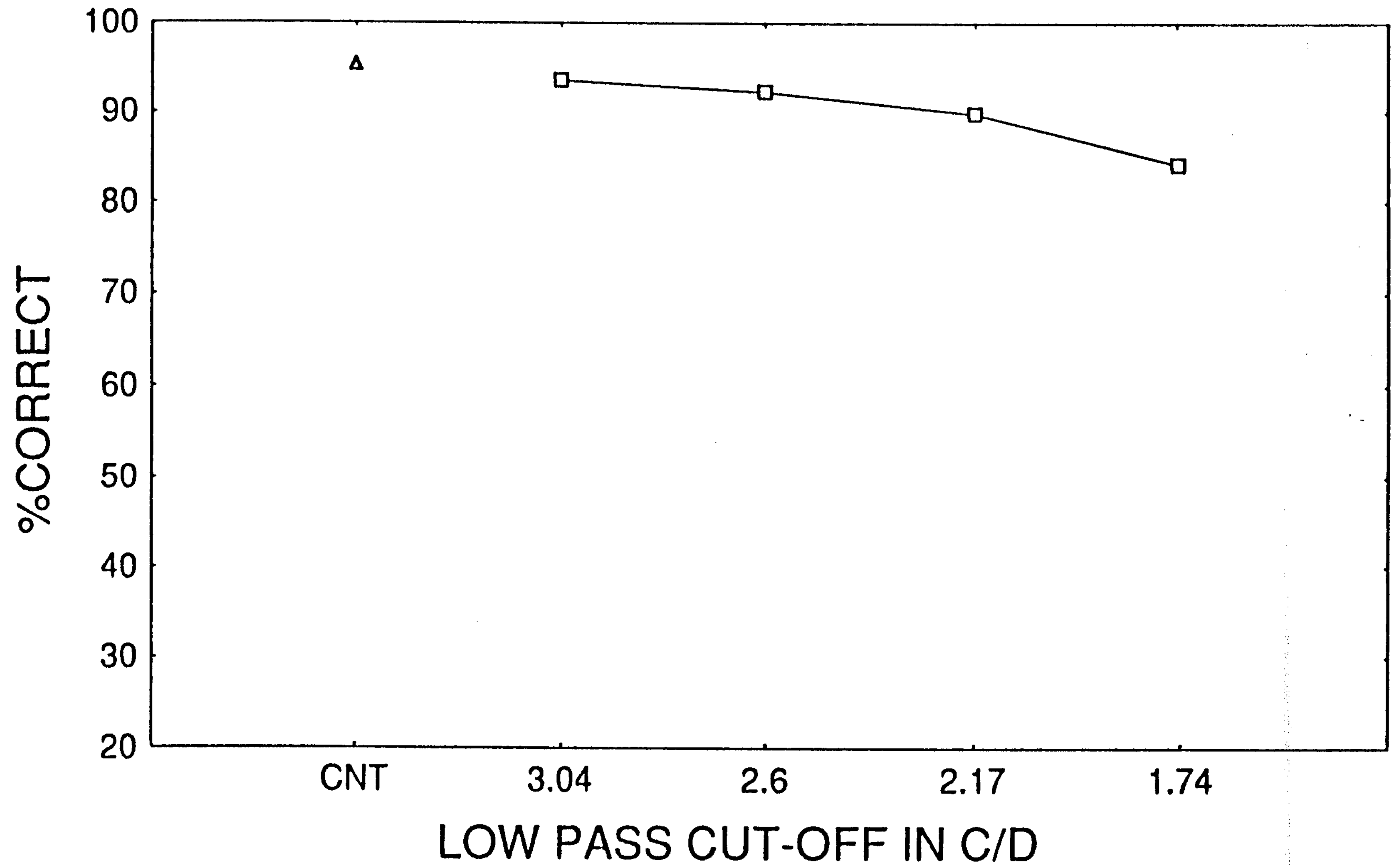

%CORRECT
100
90
80
70
60
50
40
30
20
CNT
3.04
2.6
2.17
1.74
LOW PASS CUT-OFF IN C/D

Figure 6

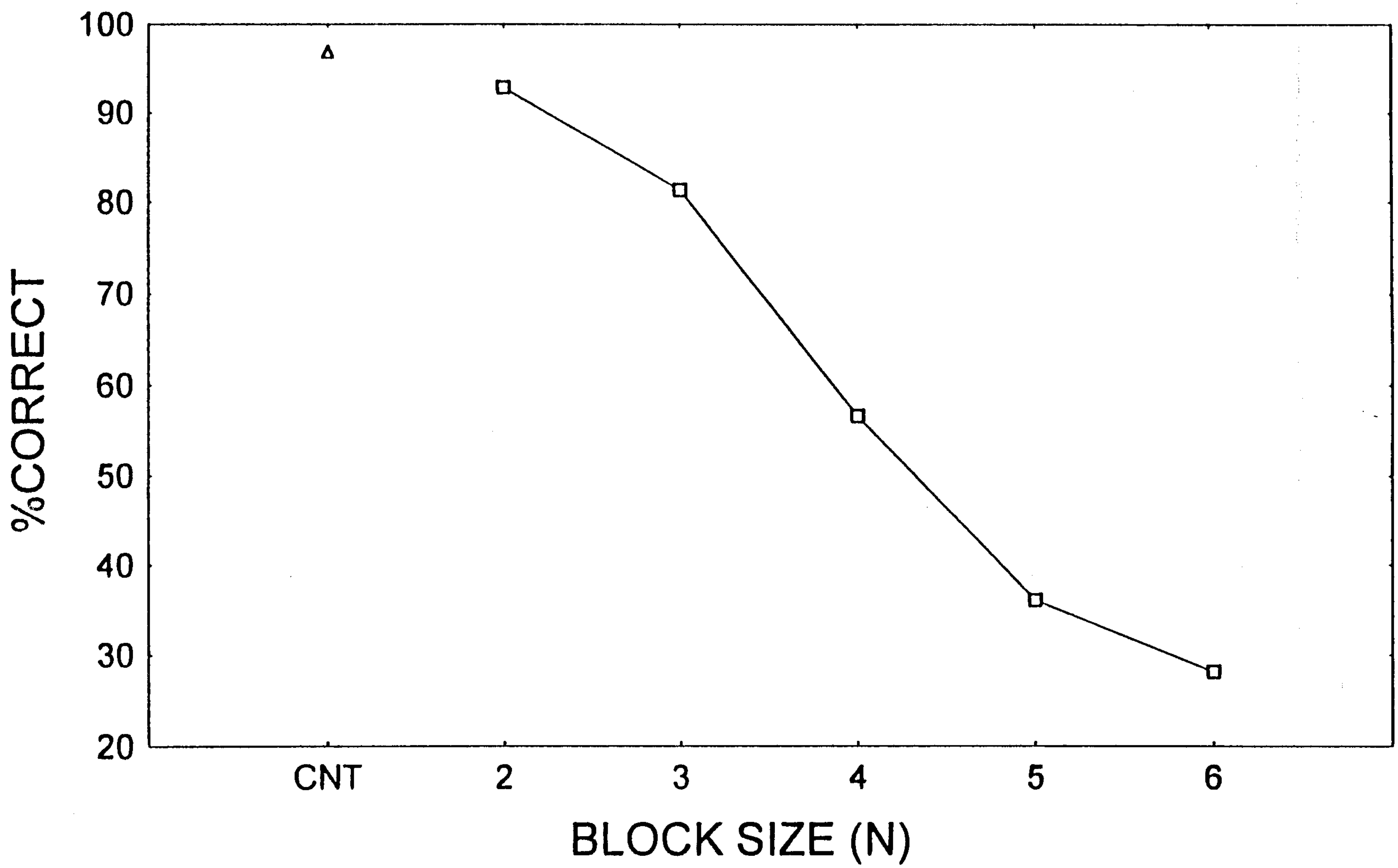
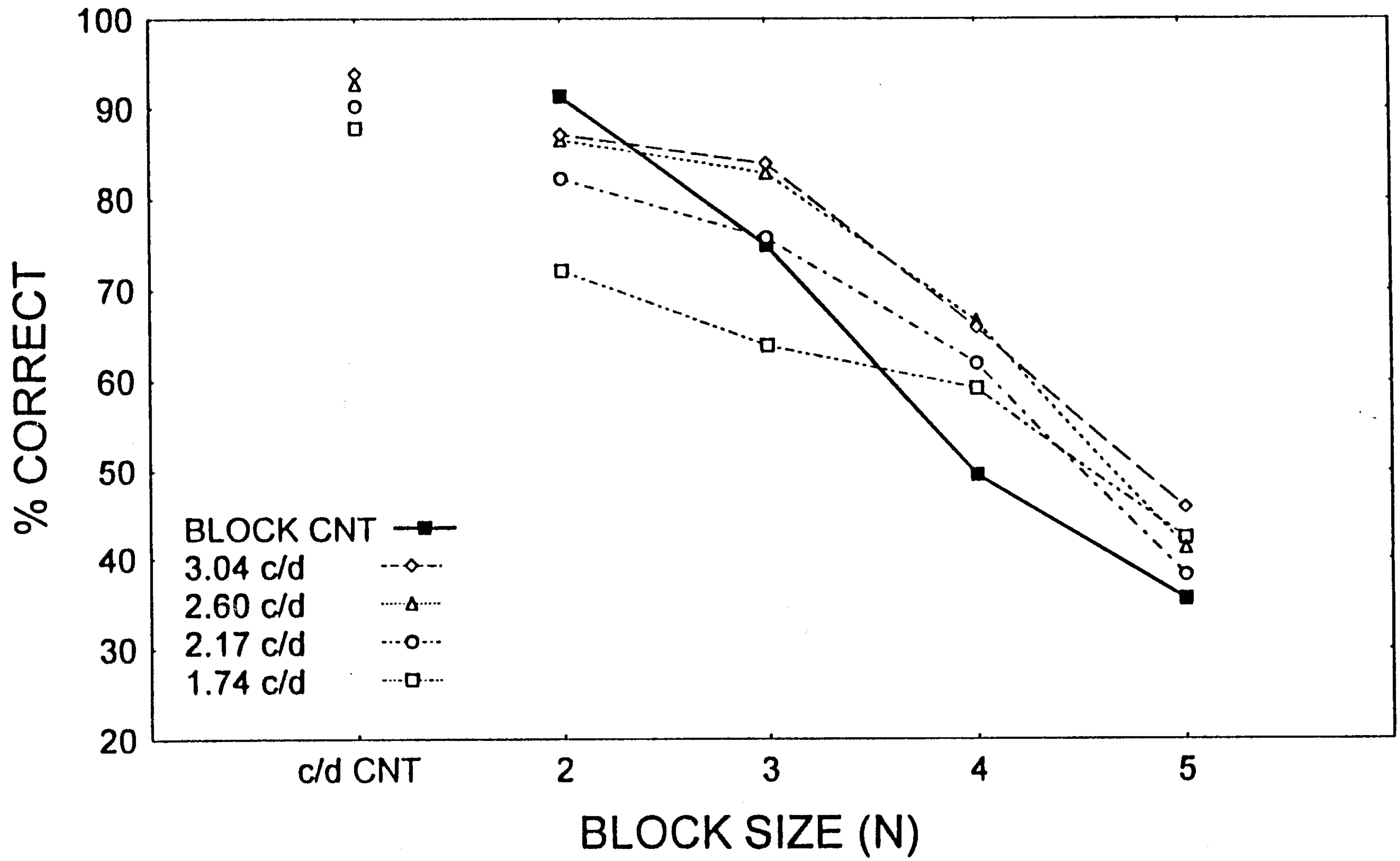


Figure 7



% CORRECT
100
90
80
70
60
50
40
30
20
BLOCK CNT
3.04 c/d
2.60 c/d
2.17 c/d
1.74 c/d
c/d CNT
2
3
4
5
BLOCK SIZE (N)

Figure 8

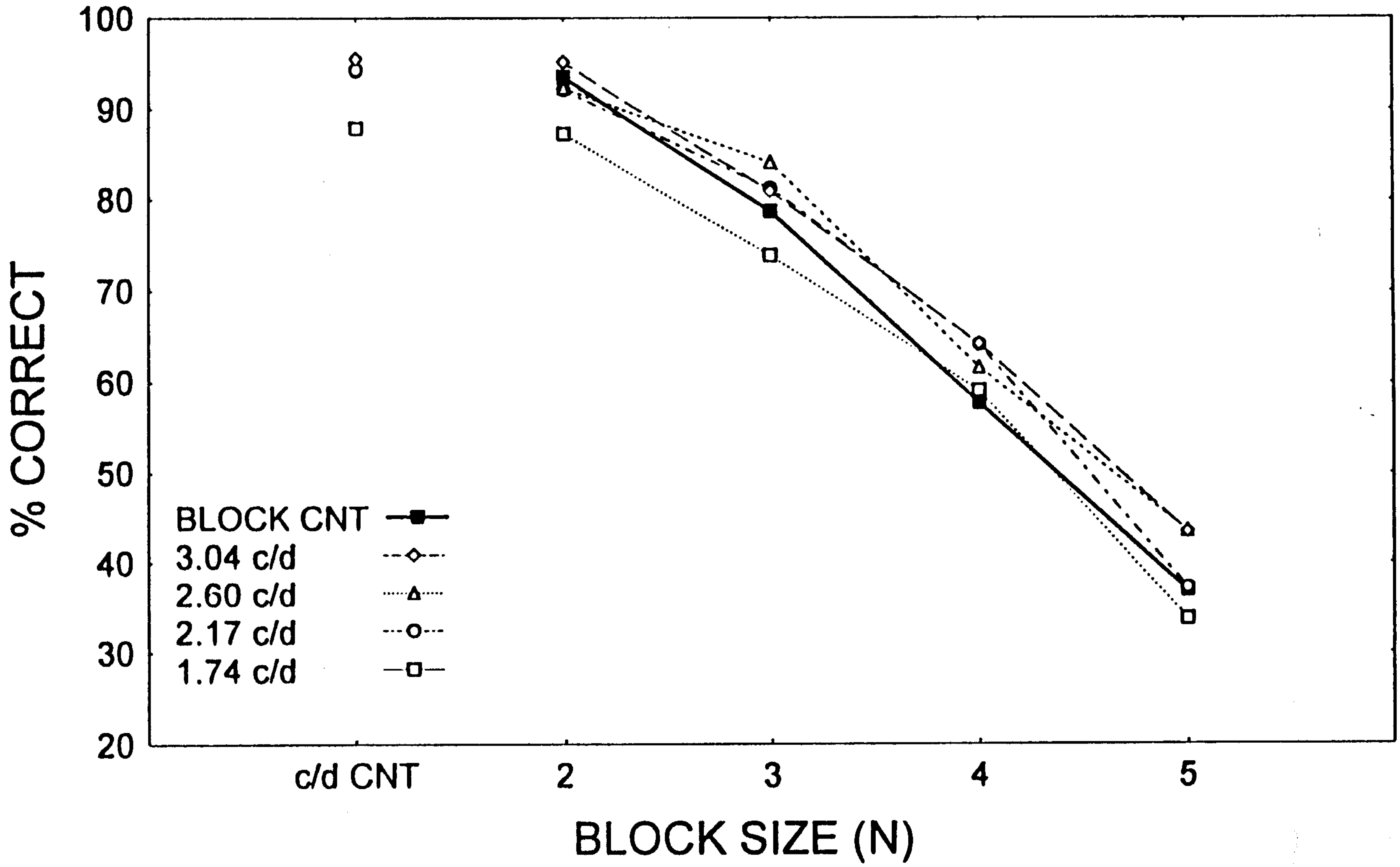
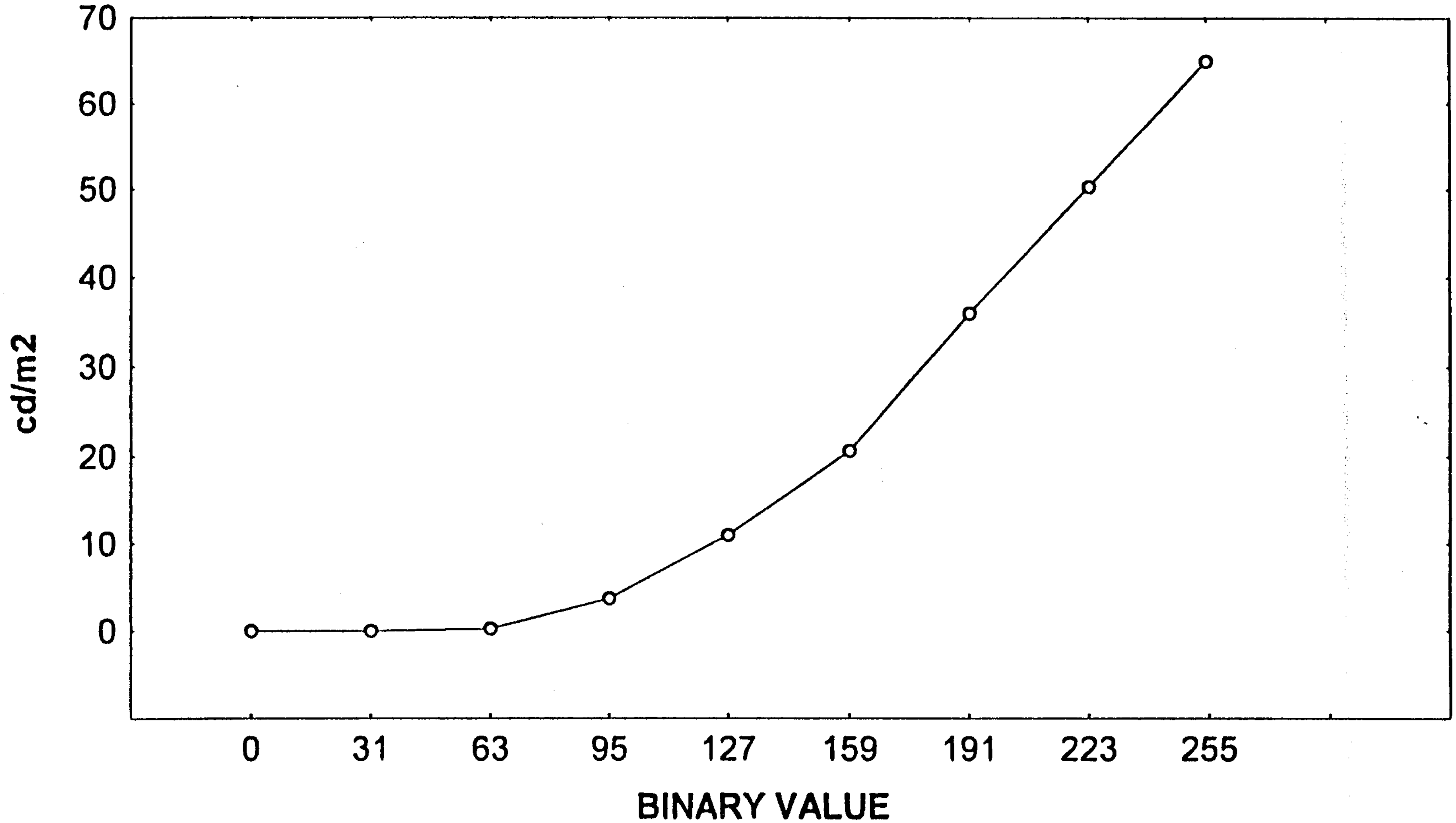

% CORRECT
100
90
80
70
60
50
40
30
20
BLOCK CNT
3.04 c/d
2.60 c/d
2.17 c/d
1.74 c/d
c/d CNT
2
3
4
5
BLOCK SIZE (N)

Figure 9



70
60
50
40
30
20
10
0
cd/m2
0
31
63
95
127
159
191
223
255
BINARY VALUE

Figure 10